

# The Pattern Basis Approach to Circuit Complexity

Bruce K. Smith\*

June 17, 2016

## Abstract

We describe and motivate a proposed new approach to the problem of finding a “complexity formula” which lower-bounds the circuit complexity (over circuits in a given class) of an arbitrary boolean function, given its truth table, and which can be used to prove nontrivial lower bounds on the complexity of specific functions of interest. This requires addressing the “natural proofs barrier” [Razborov & Rudich 1997], which (roughly) rules out any formula which can be computed in time polynomial in the truth table size. The proposed kind of formula would take as long to compute in general as a brute-force search among all possible circuits, but has a form amenable to proving lower bounds on its value for well-understood explicit functions.

This form is based on a new formalization of “patterns”, as elements of a special basis of the vector space of all truth table properties. We prove that *any* “pattern basis” gives a nonnegative and subadditive complexity formula, which (after scaling to fit primitive circuit costs) lower-bounds circuit complexity (though often trivially). We give specific conditions on the pattern basis which would lead to formula values having a useful range (in which all small circuits have low values, but most functions have high values). We discuss some of the issues involved in satisfying those conditions, and the features a pattern basis would need to make proving explicit-function lower bounds practical.

This paper is intended for all mathematicians interested in the general phenomenon of complexity, including those in other fields which might be relevant to realizing this new approach (such as linear algebra, combinatorics, and finite group representation theory).

---

\*email: oresmus@gmail.com; blog: <http://oresmus.github.io>

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
1.1	The problem of circuit complexity . . . . .	4
1.2	Structure of this paper . . . . .	7
1.3	Related work . . . . .	7
<b>2</b>	<b>Summary of approach and ideas</b>	<b>8</b>
2.1	background and terminology . . . . .	8
2.1.1	circuit models . . . . .	8
2.1.2	illustration of how a circuit is composed of generating circuits	10
2.1.3	“complexity formula” . . . . .	11
2.2	general expectations . . . . .	11
2.3	Basic idea: we want a sum of nonnegative terms... . . . .	11
2.4	... which are magnitudes of pattern match values ... . . . .	13
2.5	... which are entries in a pattern matrix ... . . . .	13
2.6	... of very high dimension ... . . . .	14
2.7	... so the log of a submultiplicative matrix measure can range up to $2^n$ . . . . .	15
2.8	This does prove lower bounds, but they’re often trivial... . . . .	15
2.9	(... even for circuits of fixed width ...) . . . . .	16
2.10	... at least when the pattern basis ignores the generating circuits. . . . .	17
2.11	We should choose it so generators’ pattern matrices are sparsest ... . .	19
2.12	... which has not yet been achieved... . . . .	20
2.13	... since naively “adjoining a wire” doubles measured complexity ... . .	21
2.14	... but there’s hope, since that’s the <i>only</i> hard issue we need to solve. . . . .	22
2.15	Proof outline: a pattern basis meeting <i>these</i> conditions will be useful. . . . .	24
2.16	desired qualitative nature of patterns . . . . .	25
2.16.1	kinds of functions . . . . .	25
2.16.2	kinds of patterns . . . . .	27
2.16.3	lower-bounding the complexity of specific functions . . . . .	30
2.17	Figures and examples . . . . .	36
2.17.1	computing measured complexity ( $C_{M,\mathcal{P}}$ ) . . . . .	36
2.17.2	examples of pattern bases . . . . .	38
2.17.3	pattern matrix for the AND function . . . . .	40
2.17.4	change of pattern basis . . . . .	42
2.17.5	comparing choices of pattern basis $\mathcal{P}$ and matrix measure $M$ . . . . .	43
<b>3</b>	<b>Notation</b>	<b>46</b>

<b>4</b>	<b>Definitions and Theorems</b>	<b>49</b>
4.1	matrix measures . . . . .	49
4.2	pattern basis (and possible generalizations) . . . . .	52
4.3	pattern matrix . . . . .	55
4.4	measured complexity functions . . . . .	57
4.4.1	unscaled version ( $C_{M,\mathcal{P}}$ ) . . . . .	58
4.4.2	scaled version ( $C_{M,\mathcal{P},\lambda}$ ) . . . . .	60
4.5	“conventional” structure of generating circuits . . . . .	61
4.6	simple case — pattern basis with ideal behavior . . . . .	62
4.7	fallbacks (for worse behavior) . . . . .	65
4.7.1	not all trivial wire operations are free . . . . .	66
4.7.2	even some wire permutations are not free . . . . .	70
4.7.3	adjoining a wire increases measured complexity . . . . .	73
<b>5</b>	<b>Motivation</b>	<b>78</b>
5.1	the “pattern intuition” . . . . .	78
5.2	can linear algebra help? . . . . .	79
5.3	higher-degree patterns . . . . .	81
5.4	pseudorandom vs random . . . . .	83
5.5	but with so many patterns, what can they be like? . . . . .	86
5.6	the property space has room for $2^{(\text{number of gates})}$ patterns . . . . .	87
<b>6</b>	<b>Discussion</b>	<b>89</b>
6.1	But does it “naturalize”? . . . . .	89
6.2	Does it “relativize” or “algebrize”? . . . . .	92
6.3	the significance of linear superpositions of pattern maps . . . . .	94
6.4	even a nonconstructive proof would be significant . . . . .	94
6.5	Conclusion . . . . .	96
	<b>Acknowledgements</b>	<b>97</b>
	<b>References</b>	<b>97</b>

# 1 Introduction

## 1.1 The problem of circuit complexity

We consider an arbitrary *boolean function*  $f$ , from  $n$  inputs to  $m$  outputs (with values in  $\mathbb{F}_2$ )<sup>1</sup>, expressed by a *truth table* (an explicit table of  $f$ 's output for each of its  $2^n$  possible inputs). We want to compute  $f$  efficiently, by expressing it as a *circuit* made of *gates* (functions from a given fixed set, each with only a few inputs and outputs), using as few gates as possible. (A *circuit* is just a composition of functions, each representing the effect of a single gate and some noninteracting “*wires*” adjoined in parallel.) The *circuit complexity* of  $f$  is the smallest “gate cost” achievable in a circuit which computes it. (More precise definitions are given in later sections. The introductory information in this section is well-known; for general references see [Wil11], [AB09], and [Aar16], and for related surveys see e.g. section 3 of [Wil14], and [All96].<sup>2</sup>)

It's well known, and easily proved by a counting argument, that most such functions  $f$  are almost maximally hard to compute, in the sense of requiring almost as many gates as the number of bits needed to specify an arbitrary truth table, which is linear in  $m$  and exponential in  $n$  (ignoring smaller factors coming from the number of available gates, and the number of ways to hook up a given sequence of gates into a circuit, which reduce the number of gates needed).

For many specific functions  $f$ , it's widely believed that they are “obviously” somewhat hard to compute, in the sense of requiring a number of gates specified by various polynomial or larger functions of  $n$ . (To be more precise, such a conjecture is about a *family*  $\langle f_n \rangle$  of functions of increasing input sizes  $n$ , and asserts (for some specific definition of circuit cost  $C_{\text{cost}}$ ) that for any family of circuits  $\langle c_n \rangle$  where each  $c_n$  computes  $f_n$ , that  $C_{\text{cost}}(c_n)$  must grow at least as fast as some given function of  $n$  (once  $n$  is sufficiently large).<sup>3</sup>)

---

<sup>1</sup> Generalizing to allow each input and output (of  $f$  or of the gates) to take values from its own finite *value-set* in place of  $\mathbb{F}_2$ , with circuit construction rules enforcing consistency of value-sets when making connections, is reasonable, and useful when considering circuit-like processes found in nature; so one test of a good theory of circuit complexity would be handling this straightforwardly. For simplicity, we won't further discuss this generalization except to indicate how it could fit into our formal framework.

<sup>2</sup> [All96] was still “depressingly up to date” in 2008, according to [All08].

<sup>3</sup> The number of outputs of  $f_n$  is typically 1, but occasionally is some other specified function of  $n$ . When it's 1, complexity theorists call the set of  $n$ -tuple inputs for which  $f_n$  outputs 1 (unioned over all  $n$ ) “the language recognized by  $\langle f_n \rangle$ ” (or by a circuit family or algorithm that computes  $\langle f_n \rangle$ ); this lets them identify “languages” with 1-output function families, since they're 1–1. The problem of computing a given 1-output function family (which either a circuit family or an algorithm

But it’s notoriously difficult to *prove* most conjectures of that form, for general boolean circuits.<sup>4</sup> (The following results relating circuit sizes to algorithmic complexity classes for “decision problems” come from the recent survey in [Wil14], section 3.)

- We can’t presently rule out that all decision problems in the complexity class  $P$  (i.e. computable by some polynomial-time algorithm in  $n$ ) — or even problems for which polynomial-time algorithms can only *verify evidence for a positive answer* (i.e. those in the class  $NP$ )<sup>5</sup> — can be computed by circuit families  $\langle c_n \rangle$  with a small *linear* size (e.g.  $C_{\text{cost}}(c_n) \leq 6n$ , for a conventional set of gates and cost function). (But most complexity theorists believe that such problems typically require circuits of polynomial or superpolynomial size, respectively.)
- For polynomial-sized circuit families (which solve the class of problems called  $P/\text{poly}$ ), we can’t rule out that they can compute all functions in certain complexity classes believed to be even larger than  $NP$ . (But most theorists believe  $P/\text{poly}$  doesn’t contain all problems in  $NP$ . The conjecture that it doesn’t implies  $P \neq NP$ , since it’s easily seen that  $P \subset P/\text{poly}$ .)

There has been gradual progress on proving limitations of more restricted circuit classes than  $P/\text{poly}$ . Even so, the huge discrepancy between what’s guessed (with much supportive evidence), vs. what’s presently provable, is widely believed by experts to indicate that new general approaches to circuit complexity are needed.

Several obstacles to proving large lower bounds on circuit complexity of explicit function families have been identified, including the “*natural proofs barrier*” of Razborov & Rudich [RR97]. This is relevant to the proof strategy of first proving the validity of a general “complexity formula” (which maps  $f$ ’s truth table to a lower bound on  $f$ ’s circuit complexity) for all boolean functions  $f$ , and then proving it has a high value for specific functions. What Razborov and Rudich showed is that, under widely believed assumptions, any such “complexity formula” would either fail

---

might solve) is also called a “decision problem”.

<sup>4</sup> There are exceptions — see e.g. [SM02], which reviews an exponential circuit size lower bound (proven in 1974) for a natural problem in logic. But nontrivial lower bounds remain unproven for *most* such conjectures, including many which “seem obvious”.

<sup>5</sup> Note that when considering circuit complexity of the “nondeterministic” class  $NP$ , the circuits themselves would still be the usual deterministic ones, unless stated otherwise (and we only consider deterministic circuits in this paper).  $NP$  is strongly suspected to be larger than  $P$  (that famous conjecture is known as  $P \neq NP$ ), and almost as strongly suspected to contain problems which take exponential time (for worst case inputs) to compute by a deterministic algorithm (that’s the “Exponential Time Hypothesis” [IP01]).

to prove high complexity for all but a tiny fraction of functions (even though most functions *have* high complexity), or be itself hard to compute, in the sense of taking time superpolynomial in the size of  $f$ 's truth table data — unlike all functions which had been previously used in related proofs. (Note that the *actual* (exact) complexity function can be computed by the simple brute force algorithm of considering all circuits (in order of increasing cost) until a correct one is found, but for most  $f$  this takes exponential time in the truth table size.)

The present approach takes the view that it's only to be expected (in hindsight) that a generally applicable complexity formula would be hard to compute — the phenomenon of complexity is, after all, complicated — but that finding a complexity formula applicable to most functions is nevertheless essential, at least if its validity is to be proven inductively over function composition, for well-known reasons discussed later — basically that any function  $f$  can be expressed as  $g \circ h$  where both  $g$  and  $h$  are individually almost random.<sup>6</sup> We also hold that being hard to compute in general doesn't rule out making practical use of such a formula for sufficiently well-understood specific functions; we'll elaborate on this in later sections.<sup>7</sup>

With respect to looking for a complexity formula which gives nontrivial lower bounds for most functions, this state of affairs begs several questions:

1. What kind of complexity formula might be (a) hard enough to compute from the truth table to not be ruled out by the “natural proofs barrier”, yet also (b) “make sense”, both in terms of our intuitive reasons for believing certain specific functions to be hard, and in being provably correct as a complexity lower bound? (And, of course, (c) how would its bounding property actually be proved?)
2. If such a formula was found, how might it be used in practice to prove specific interesting functions complex? (Ideally we want to prove them almost as complex as we suspect they are, but the current state of our knowledge is so poor that even proving some problem in **NP** had *superlinear* circuit complexity would be a breakthrough. [Wil14])
3. How might such a formula distinguish “random” functions (which have high complexity) from “pseudorandom” functions (which have low complexity, but

---

<sup>6</sup> A similar issue for formula complexity, related to expressing  $f$  as  $g + h$ , is discussed in [RR97].

<sup>7</sup> This is similar to the “personal view” expressed in [AB09] of how best to react to the natural proofs barrier. They cite an example from mathematics of a graph-theoretic lower bound which is very slow to compute in the general case, but can be evaluated exactly for a special case of interest. I have also seen similar views expressed by several other people.

“look random” to all practical tests)? (All functions we consider in this paper are deterministic, but this question makes sense if we interpret the term “random functions” to mean “most functions”, or “functions of near-maximal circuit complexity”, or “functions of near-maximal Kolmogorov complexity”.)

Answering those questions would be highly useful, but for a truly general theory we’d also want to know:

4. How might such a formula be customized for different sets or costs of primitive gates, which in some cases would compute only functions in smaller families than the general boolean functions described above, e.g. reversible or  $\mathbb{F}_2$ -linear functions?

The purpose of this paper is to outline a new proposal for approaching this problem and answering those questions, both at the informal level of basic intuition and motivation, and as a formal framework which suggests specific forms of potential complexity function to investigate, and provides reason to hope some of them might have the required properties.

## 1.2 Structure of this paper

In Section 2 we summarize the proposed framework and the ideas involved in it, and explain our anticipated answers to the questions listed above. We give explicit properties which a “pattern basis” would need (relative to a given class of allowed circuits and their cost function) to lead to a useful complexity formula. We present some illustrative examples and figures in Section 2.17.

In Sections 3 and 4 we describe the proposed framework formally, mention some simple generalizations, and prove its basic properties, including the sufficiency of the “conditions for usefulness” given in Section 2.

In Section 5 we motivate our approach at an intuitive level, giving lines of argument which seem to lead to something like the proposed framework. In Section 6 we address whether this approach might “naturalize”, “relativize”, or “algebrize”, further discuss some of its mathematical features, and give a brief conclusion.

## 1.3 Related work

As far as I know, no other proposal for a possible kind of “complexity formula” (for lowerbounding the complexity of any boolean function, given its truth table) directly addresses the basic questions listed in Section 1.1 — particularly how it might be

hard enough to evaluate in general to evade the natural proofs barrier (while still proving most functions complex), but at the same time be potentially usable in proofs about explicit functions; and also how its lowerbounding property might be provable. (For overviews of approaches to the related  $P$  vs.  $NP$  problem, see [For09] and [Aar16]; for recent surveys of circuit complexity lower bounds, see the references in Section 1.1.)

The Geometric Complexity Theory research program [MS01] might be able to circumvent the natural proofs barrier (see also a different author’s “lay overview” [Reg02], and [Aar16]). Its basic approach is to exploit symmetries in very special problems [Mul12]; as far as I can tell, it doesn’t try to be applicable in principle to *all* functions, at least not by proposing a lower bound formula which is “subadditive with respect to function composition”. (My understanding of GCT and its background is quite limited, so I can’t rule out its containing related ideas implicitly.)

Like the present proposal (which makes use of a “property space” of doubly-exponential size, in the number of inputs of a function being analyzed), GCT involves very large vector spaces; but this may be only a superficial similarity, since it seems to use them in an unrelated way. Likewise, group representations are a central topic in GCT, and this paper speculates that finite group representation theory might help in realizing the present proposal (see Section 4.2); but both the relevant groups, and how they might be used, seem to be different.

(Note that the present proposal’s definition of “pattern matrix” is not related to the “pattern matrix method for communication lower bounds” [She08], or to the “pattern matrix” in factor analysis [WikF].)

## 2 Summary of approach and ideas

### 2.1 background and terminology

#### 2.1.1 circuit models

Given a specific *circuit model* (set of *primitive gates*, *construction rules* for combining them into circuits, and definition of nonnegative *circuit cost* which is additive when circuits are serially composed), the *circuit complexity* of a function is defined as the minimum cost of any constructible circuit which computes it (or infinity, if no constructible circuit computes it).

We don’t require that a circuit model can construct circuits which compute every function, nor that it treat common simple operations (like interchanging the order



of wires) as having zero cost (or even as always being permitted), since we want the general framework to apply to intentionally limited models, such as models which can construct only:

- reversible functions — e.g. all primitive gates are small (reversible) “*S-boxes*”<sup>8</sup>, and wires can’t be split or discarded (typically this can construct any even permutation of the  $2^n$  input vectors [CG75]); or
- $\mathbb{F}_2$ -linear functions — e.g. wires *can* be split or discarded, but only XOR gates (which do addition in  $\mathbb{F}_2$ ) can be used.

But our main focus in this paper is on a conventional model for “*general boolean computation*”, whose circuits can compute any boolean function (with any number of inputs and outputs), and which treats permuting or splitting wires as free of cost. (Such models are well-known, and the choice among them has only a linear effect on circuit complexity [Wil11], so we have no need in this paper to pick a specific one.)

Without loss of generality, we assume each circuit model also specifies a *generating set*  $\mathcal{G}$  of *generating circuits*, which are sufficient (when serially composed) to generate the same constructible functions without imposing a cost penalty — that is, the minimum cost of computing any function using a serial composition of generating circuits is the same as the minimum over all constructible circuits in the model.<sup>9</sup> (The set  $\mathcal{G}$  will be infinite in models permitting arbitrarily high “circuit widths”, but would typically have a simple structure; see Section 4.5 for a conventional construction of  $\mathcal{G}$ .)

Without this assumption, it would be harder to analyze circuit cost when the model’s construction rules include “*parallel composition*”, i.e. placing non-interacting circuits side by side to make one larger circuit, combining their respective sets of inputs and outputs by disjoint union. We don’t want to require every model to simply add costs in that case (though typical models would do that); by making the model predefine an adequate set of generating circuits, which need only serial composition to be further combined, we avoid that issue.

This assumption is reasonable in practice — in a typical model, the generating circuits would include any circuit containing at most one primitive gate or trivial

---

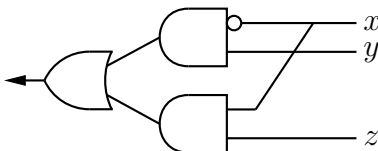
<sup>8</sup> In this paper, by “S-box” we mean a reversible boolean function on a small number of bits (or a gate which computes such a function); by “reversible” we mean bijective (invertible and onto).

<sup>9</sup> Note that in spite of this assumption, there is no guarantee that a generating circuit is the lowest-cost way to compute its own function — some composition of other generating circuits might in principle have lower cost.

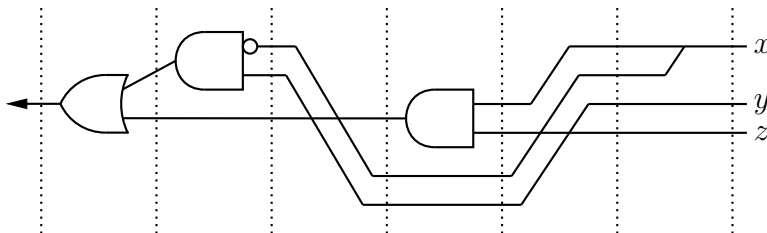
wire operation (like “splitting a wire”<sup>10</sup>), alongside any number of non-interacting parallel wires (possibly permuted); then any circuit could simply be *viewed* as a serial composition of generating circuits, without affecting its cost. Accordingly, from now on “**composition of circuits**” always means “serial composition”, which also composes the functions computed by those circuits.

### 2.1.2 illustration of how a circuit is composed of generating circuits

**Figure 1:** A simple multiplexer circuit, shown in a conventional form.



**Figure 2:** The same circuit as in Figure 1, shown as a composition of 6 generating circuits (which have the “conventional” structure described in Section 4.5).



The point of this circuit representation is that any function computed by a whole circuit is a composition of functions computed by only a few types of generating circuit. The generating circuits used here include 3 gates, 2 wire permutations, and a “splitter”, in some cases with 1 or 2 “adjoined wires” (shown below the gate or splitter, since they correspond to their generating circuit’s highest-numbered inputs and outputs, but logically running “alongside” or “in parallel to” the gate or splitter).

The “widest” generating circuits in this example have 4 inputs and 3 or 4 outputs: the AND gate with 2 adjoined wires, and the permutation of 4 wires. (The circuit as a whole has 3 inputs and 1 output, just as in Figure 1.)

<sup>10</sup> In most formalizations of circuits, all gates have exactly one output, which can be “split” into multiple wires by giving the gate a “fanout” greater than 1. In our formalization, this would be confusing (since gates are allowed to have multiple outputs with different values); for that and other reasons, we replace the concept of “fanout” by the operation of “splitting a wire”, which doesn’t directly involve any gates (though it could alternatively be thought of as a gate with one input and two outputs, all with equal values). Similarly, by “discarding a wire” we mean letting it “end” within the circuit; this could be thought of as a gate with one input and no outputs.

### 2.1.3 “complexity formula”

By a ***complexity formula*** (for a specific circuit model), we mean a formula which can be applied to the truth table of an arbitrary function  $f$ , which always produces a lower bound on its actual circuit complexity (for that circuit model). The function computed by such a formula can be called a ***complexity measure***, or “a kind of ***measured complexity***” when we want to emphasize that it’s only a lower bound, not necessarily a good one. We are just as concerned with the formula itself (an expression of a certain form) as with the function it computes, since we need to ensure it’s possible to prove lower bounds on its values for certain kinds of functions.

We informally call a complexity formula ***useful*** if it produces reasonably good lower bounds, at least for some functions of special interest and for most arbitrary functions. (Note that most functions have near-maximal complexity. The requirement of being useful for most functions might not matter for some applications, but for an approach like this one (which would prove its lower bound inductively over function composition), we believe it’s necessary, for well-known reasons we’ll discuss later.)

In this paper, by ***complexity*** alone we mean either circuit or measured complexity, according to context. (When other kinds of complexity are needed, we’ll name them explicitly.)

## 2.2 general expectations

We expect that for a given circuit model, there may be more than one kind of useful measured complexity, and probably no single formula (of a useful form) gives an exact measure of complexity for all functions.

Even a single example of a useful complexity formula might be *very* useful. But given these expectations, and guessing there might be no canonical choice of useful complexity formula, we’re interested in techniques for coming up with families of (provably valid) complexity formulae which are candidates for being useful. (This paper presents one such technique.)

## 2.3 Basic idea: we want a sum of nonnegative terms...

Now we turn to the basic ideas behind our proposed approach. To start with, we’ll describe it “from the outside in”, so we can discuss how it might evade the natural proofs barrier before anything else. The remainder of Section 2 will fill in the details, including what’s still required to make the framework useful. (The figures in Section 2.17 illustrate how the proposed kind of formula would be computed for some

simple examples, and compare the effects of several choices of pattern basis. See also Section 5 (Motivation) for a presentation from a completely different direction.)

The proposed formula for a complexity lower bound is analogous to “the log of the sum of a very large number of nonnegative terms” — so many terms that computing the log of the sum directly (from an arbitrary truth table) takes too long to be ruled out by the natural proofs barrier; but with all terms nonnegative, so lower bounds on the formula’s value for specific functions can in principle be proven, if enough is known about those functions to prove many specific terms are positive. (The exact nature of the formula, and of the “sum of terms” it contains, will be discussed below. See Section 6.1 for a more technical discussion of why the natural proofs barrier seems unlikely to rule out the proposed kind of formula.)

One of our goals is to construct the formula so that, for any function  $f$ , most of the summed terms are zero (or low-valued enough to be insignificant), and (for most  $f$ ) the relatively few nonzero terms are “difficult to find” (in spite of their absolute number being large). This means you can’t even *estimate*  $f$ ’s complexity by random probing of only a polynomial number of terms (relative to the size of  $f$ ’s truth table), since it’s likely all the terms you probe will be zero<sup>11</sup> — unless you know *which* terms to probe (since you know or guess something about  $f$ ’s structure), or unless you probe all the “simple” terms and some of *those* are nonzero (which is often true about special or interesting  $f$ , but isn’t true for most  $f$ ). This is important, since otherwise a randomized polynomial algorithm could estimate measured complexity well enough to distinguish low-complexity functions from random functions, which would give us a “useful, large, and constructive” property (in the terminology of the natural proofs barrier) — that is, it would prove (subject to widely believed assumptions about pseudorandomness) that our approach can’t be made to work.

As mentioned above, the terms in our formula’s sum are always nonnegative — this lets you prove lower bounds on the sum (and thus on its log, and thus on  $f$ ’s complexity) if you know the “location” (i.e. the ***term index***, or position in the expression for the sum) of enough positive terms — which you might know, if you understand the structure of  $f$ , for example for explicit functions  $f$ . (Furthermore, if you don’t need an *optimal* bound, it’s sufficient to lower-bound only *some* of the positive terms, which means partial knowledge of  $f$ ’s structure or of  $f$  itself might still allow you to prove a useful lower bound on its complexity.)

---

<sup>11</sup> Except possibly for a few “non-discriminating” terms (which are each nonzero for *many* functions), which we can ignore here.

## 2.4 ... which are magnitudes of pattern match values ...

The terms in the “sum” are the nonnegative “*magnitudes*” of *pattern match values* (or *pattern values* for short), which are numbers (possibly negative or complex) in some field  $K$  (probably  $\mathbb{C}$  or  $\mathbb{R}$ , or possibly a finite field). (The magnitude measure  $M: K \rightarrow \mathbb{R}^{\geq 0}$  (on pattern values) could be any submultiplicative and subadditive function, such as 0 for 0 and 1 for all other values, any submultiplicative norm over  $K$ , or other possibilities. Generalizing  $K$  to a perhaps-noncommutative ring might be worth considering, but is not further discussed here, except to note that many of our definitions and proofs would work without change in that case.)

The intuitive interpretation is that each term measures the degree to which  $f$ ’s truth table “matches a specific pattern”; the pattern value (and therefore the term which measures its magnitude, which we’ll also call a *pattern magnitude*) can be computed from the truth table given the term index. (Note that for most patterns, the functions to compute their values would be very complex, and whatever was special about the matching truth tables would not be something “visible” or “obvious”. It’s no coincidence that the patterns we usually “see” or measure are the “simple” ones, which are a small subset of all patterns, as discussed further below.)

The pattern value itself (as opposed to its magnitude) can be a negative or complex number, encoding details about *how* the pattern is matched, so it can “interfere” with other pattern values when two boolean functions are composed and their pattern values are multiplied, as described in the next section.

## 2.5 ... which are entries in a pattern matrix ...

The pattern values are the entries of a very large *pattern matrix*, defined so that composing functions corresponds to multiplying these matrices; this lets us prove our measured complexity formula is a lower bound on the true circuit complexity, by proving it’s a “subadditive matrix measure” (at least for matrices of this special form), and that it’s correct for the circuit model’s generating circuits (e.g. all circuits containing at most one nontrivial gate, alongside any number of parallel wires).

By a “*subadditive matrix measure*” in this context, we mean any  $C$  which takes pattern matrices to values in  $\mathbb{R}^{\geq 0}$  and for which  $C(AB) \leq C(A) + C(B)$ . If such a  $C$  is applied to a circuit or circuit function, we treat it as being applied to the associated pattern matrix; in that case we’d have  $C(f \circ g) \leq C(f) + C(g)$ . Calling this “subadditive” is an abuse of terminology, justified by thinking of composition of circuits as being like addition, at least when circuit cost or complexity is being

discussed.<sup>12</sup>

One possibility for the choice of subadditive matrix measure (which we'll assume is our actual choice, in most of what follows) is simply the log of the “sum of pattern magnitudes”, with the “sum” itself actually being a “*submultiplicative matrix measure*” with value at least 1; such a measure might be as simple as the maximum number of nonzero entries in any row (and for simplicity we'll assume it's exactly that, for most of our discussion). (By a ***submultiplicative matrix measure*** we mean any  $M$  which takes pattern matrices to values in  $\mathbb{R}^{\geq 0}$  and for which  $M(AB) \leq M(A)M(B)$ . This is a standard use of the term “submultiplicative” (except that we only require it to work for pattern matrices, not necessarily for all matrices). For more examples and related theorems, see Section 4.)

(To clarify: the phrase “sum of pattern magnitudes” is meant to be suggestive, with “submultiplicative matrix measure applied to matrix of pattern values” being its precise version — even though such a matrix measure would not just be a “sum of all the magnitudes”. It might be the maximum over rows of the sum of magnitudes in each row, or something more complicated; but like an actual sum (of nonnegative values), it could be lowerbounded by lowerbounding subsets of the terms.)

## 2.6 ... of very high dimension ...

The pattern matrix dimensions for an arbitrary boolean function  $f$  from  $n$  bits to  $m$  bits are not  $m$  by  $n$ , or even  $2^m$  by  $2^n$  (like the matrices used to describe quantum computation, when  $m = n$ ), but  $2^{2^m}$  by  $2^{2^n}$ , since the matrix represents a linear map between arbitrary  $K$ -valued ***properties*** (functions) of *truth tables* (not just properties of bit vectors).<sup>13</sup>

Specifically (as demonstrated in Section 2.17.3, and described formally in Section 4.3), we reinterpret  $f$  as mapping truth tables (presented to it on its output, describing arbitrary boolean functions from  $m$  bits to 1 bit) into other truth tables (visible on its input to whatever uses it, thus describing boolean functions from  $n$  bits to 1 bit, created by composition of (the function computed by) the  $m$ -bit truth table with  $f$ ). We then “linearize” this (so it maps formal linear combinations of truth tables (with coefficients in  $K$ ) from  $f$ 's outputs to its inputs), and finally dualize it to get  $f$ 's ***pattern map***, a linear map from  $f$ 's inputs' ***property space*** (vector

---

<sup>12</sup> In this paper we never need the conventional meaning of “subadditive” for a measure on matrices or functions. In some other context where we did, we might distinguish these terms by referring to this paper's version as “p-subadditive” or “product-subadditive”. But for measures on the field  $K$ , we *do* use “subadditive” with its conventional meaning.

<sup>13</sup> This is the simplest of several possibilities, but the others would also involve large dimensions (see Section 4.2).

space of arbitrary  $K$ -valued functions of truth tables with 1 output and  $n$  inputs) to its outputs' property space (the same, but over truth tables with  $m$  inputs). These vector spaces (over the field  $K$ ) have dimensions  $2^{2^n}$  and  $2^{2^m}$  respectively, so we represent the pattern map by a  $2^{2^m}$  by  $2^{2^n}$  pattern matrix (with entries in  $K$ ), as said above — but using a specially chosen ***pattern basis*** in each property space, discussed below, to make the matrix entries appropriate for the complexity formula.

(Note that we can't assume the pattern basis will be orthogonal, relative to the standard ***indicator basis*** (whose elements are properties valued at 1 for one truth table and 0 for all others). If we refer to a specific correspondence between a pattern and its ***“dual pattern”*** (a formal linear combination of truth tables), we mean the corresponding element in the dual basis of the pattern basis.)

## 2.7 ... so the log of a submultiplicative matrix measure can range up to $2^n$ .

A submultiplicative matrix measure (like the one which counts nonzero entries in rows) can be made to have values that range from 1 to (approximately) one of the matrix dimensions, so the large size of the pattern matrix allows the log of such a measure to be roughly as high as the maximum possible circuit complexity of a function, which (for circuit models which permit general boolean functions, and have not too many generating circuits, each with small cost) is comparable to the number of bits in a truth table,  $m2^n$ . (The fact that a complexity measured this way could not get *quite* that high, especially since the pattern matrices should be sparse (as we'll discuss below), is significant, but not a problem for some important applications; later we'll discuss what that fact might mean, as well as giving more principled reasons for this formal setup than the size range of values it can produce.)

## 2.8 This does prove lower bounds, but they're often trivial...

If we choose any field  $K$ , and define a pattern basis as *any* fixed choice of basis for each size of property space, then for all choices of pattern basis and many choices of submultiplicative matrix measure, we could construct a “subadditive” function of  $f$ 's truth table in the sense described above, and prove that it lower-bounded any  $f$ 's circuit complexity (after scaling it to fit the costs of the generating circuits of whatever numbers of inputs and outputs  $(n, m)$  might be needed).

But for most (i.e. for “random”) choices of pattern basis, this result would be useless, for two reasons:

- the **arbitrary width issue**: We don’t know in advance how “wide” a circuit might be needed to compute  $f$  (where for any circuit expressed as a composition of generating circuits  $g_i$ , called its **circuit stages**, by **circuit width** we mean the maximum  $n_{g_i}$  or  $m_{g_i}$  of any circuit stage  $g_i: \mathbb{F}_2^{n_{g_i}} \rightarrow \mathbb{F}_2^{m_{g_i}}$ ); but the ratio of (unscaled) measured complexity to actual circuit cost might grow arbitrarily high as a generating circuit is “widened” by adjoining non-interacting wires, as our pure-serial-composition model requires us to do. So to scale the measured complexity correctly for all possible widths, the scaling factor would often have to be 0. (Technically, we could limit the width to the maximum possible circuit complexity for the whole function  $f$ , but in practice this would be just as bad.)
- the **arbitrary basis issue**: Even for a fixed width  $w$ , if the pattern basis is arbitrary relative to the generating circuits, it will probably treat their functions like any other functions of the same **size**  $(n, m) = (w, w)$ ; thus the resulting complexity measure will probably consider the generating circuits as having almost maximal complexity for their size (as it does for most functions), rather than as having especially low complexity (as we want it to do).

These reasons turn out to be related — addressing the “arbitrary basis issue”, well enough to make a useful pattern basis for *fixed* widths, will also solve the “arbitrary width issue”. (To get ahead of the story, the basic reason is that they must both be addressed by making sure that trivial width-increasing operations like “adjoining a non-interacting parallel wire” don’t increase a circuit’s measured complexity, at least not by too much. We’ll cover this in detail later on.)

But for now, the situation will be clearer if we focus on the special case where all generating circuits are constrained to have the same width  $w$ , so  $n = m = w$  for all functions we’re considering ( $f$  and each  $g_i$ ). (This is natural for reversible circuits, but doesn’t imply reversibility.)

## 2.9 (... even for circuits of fixed width ...)

In the context of fixed-width circuits ( $n = m = w$  for all stages, for some constant  $w$ ), all we’re saying about an arbitrary “random” pattern basis is that there’s nothing special about a generating circuit’s function compared to any other non-identity function, and most functions will have near-maximal measured complexity relative to that basis (since their pattern matrices won’t happen to be sparse), so the generating circuits probably will too. (Since the measured complexity is easily proven “subadditive”, we could still scale it to fit the generating circuits’ costs (unless some



have 0 cost — we’ll discuss those later), but this would just result in a trivial lower bound which proves that computing  $f$  requires at least one generating circuit.)

So does *any* basis (of the property space for width  $w$ ) avoid seeing the generating circuits as highly complex?

For reversible functions, at least two very different bases do avoid it, but as pattern bases they’re “degenerate” — so symmetrical that they give *every* reversible function a measured complexity of 0. These are the ***indicator basis*** (in which each element recognizes a single truth table), and the ***monomial basis*** (each element evaluates the XOR of a fixed but arbitrary subset of truth table entries, representing the result in  $\{1, -1\}$ ; these bases are discussed further in later sections). Though these are as far apart as any two bases can be (according to the standard inner product in the property space), both of them are just *permuted* by every reversible function, which is why they measure its complexity as 0. (Even irreversible functions have simple actions on these bases, but in different ways — the indicator basis measures *every* function’s “complexity” as 0, but the monomial basis measures positive “complexity” in irreversible functions which “forget” information about their input (by mapping multiple inputs in  $\mathbb{F}_2^n$  to the same output in  $\mathbb{F}_2^m$ ). Note that this means there is a “subadditive” measure related to “forgetting”, but doesn’t imply it has anything to do with actual circuit complexity.<sup>14</sup>)

So it’s easy to find a basis which sees most functions as high complexity (including the generating circuits), or one which sees all reversible functions as 0 complexity, or even one which sees *all* functions as 0 complexity. But if we want a *useful* basis, we’ll need something more sophisticated.

## 2.10 ... at least when the pattern basis ignores the generating circuits.

None of the example bases considered so far depended at all on the generating circuits — so it’s not surprising their treatment of those circuits was trivial or generic. Before addressing how a pattern basis *should* take them into account, it’s worth pointing out one reason we know it will have to — the set of generating circuits may seem special to *us*, but to the property space they’re more or less “just another set

---

<sup>14</sup> This measure is shown in Table 2 of Section 2.17.5 as  $C_{M,\mathcal{P}}$  for  $M = M_{\text{nz}}$  and  $\mathcal{P} = \mathcal{P}_{\text{mon}}$ . What it actually measures is how much smaller a set of possible inputs (a subset of  $\mathbb{F}_2^n$ ) can become, after  $f$  maps it into some subset of  $\mathbb{F}_2^m$ . (This is of course maximized by the entire input set, but expressing it as a maximum over subsets makes the nature of its subadditivity clearer.) Incidentally, this is an example of a natural “subadditive” measure definable by this framework, whose value can range up to almost  $2^n$ .

of non-identity functions”.

To be more specific: there are many automorphisms on the set of all boolean functions from  $\mathbb{F}_2^n$  to  $\mathbb{F}_2^m$ , corresponding to permutations of their  $2^n$  input values and of their  $2^m$  output values. These automorphisms map the functions computed by the generating circuits to almost-arbitrary other functions. This means that even if we happen upon a “good” pattern basis in some intrinsic sense, it might be measuring a different kind of complexity than we’re interested in. For example:

- It might be measuring complexity relative to some arbitrary encryption scheme for the tuples of bits passing from each circuit stage to the next. Such a formula would claim correctly that most but not all functions have near-maximal complexity, but would disagree with us about which functions to consider simple. But it would do this consistently, in the sense of being “subadditive” over function composition. The functions it considered as having minimal complexity, though appearing to us as individually almost random, would obey the same relations among themselves as a more “conventional” set of generating circuits would — for example, typically each of them would commute with many of the others.
- Or it might be measuring complexity for a different primitive gate set, or using different gate costs, or even for computation of a different nature. (Arbitrary maps from output to input truth tables (not to mention linear maps between formal  $\mathbb{C}$ -linear combinations of them) can represent some kinds of computation that aren’t physically possible (and perhaps some that are physically possible, but not “classical”), which nonetheless could conceivably have a good measured complexity function which this setup could represent.)

So far we’ve seen ways to choose the pattern basis to get complexity lower bounds that are technically correct (after scaling), but trivial, and only in extreme ways — most functions high complexity, or all *reversible* functions 0 complexity, or *all* functions 0 complexity. What we need is something in between those extremes. But it needs to be “in between” in just the right way — not only measuring complexities with a high “dynamic range”, but with the functions computed by the *generating circuits* being the ones it deems simplest — and for widening those functions (by adding new non-interacting inputs, and corresponding outputs, to represent adjoined parallel wires) not to increase their perceived complexity.

It’s time to examine more carefully exactly what that would require, in terms of specific features of the pattern basis which we might be able to understand how to achieve.

## 2.11 We should choose it so generators’ pattern matrices are sparsest ...

Technically, by a *pattern* we just mean “an element of the chosen pattern basis” (in any property space we’re using). But to make the pattern basis *useful*, we have to choose it correctly, which means taking account of the generating circuits, so the measured complexity can have low (but mostly nonzero) values for their functions, but high values for most functions.

This means the patterns in the basis should not just be *any* properties, but properties with special significance (when they match the truth table of some 1-output function  $f$ ) for predicting the values of other patterns in the composition of  $f$  with a generating circuit. Or to be more precise, we want pattern matrices (when expressed in the chosen pattern basis) to be as sparse as possible — even for arbitrary functions (so most of their pattern values are 0, for the reasons given earlier), but especially for simple (low-complexity) functions (so they have smaller matrix measures). For the functions we want to consider simplest of all, but nontrivial — i.e., the functions computed directly by generating circuits — we want pattern matrices with only a few nonzero entries per row (or only a few “large” entries, if we’re using a matrix measure which can effectively ignore large numbers of “small” entries) — but not just 1 nonzero entry on every row (for every generating circuit), or they (and thus every constructible function) will end up with a measured complexity of 0.

Note that it’s ok if *some* generating circuits (which we want to consider trivial), e.g. those which only permute the order of wires, have only 1 nonzero pattern value per row, and thus a measured complexity of 0, as long as those circuits alone can’t be composed to make too many different functions — since their compositions will also have measured complexity of 0. (When those functions are reversible, we can achieve this by making sure the pattern basis is symmetric relative to those functions, so they just permute it (and perhaps also multiply its elements by roots of unity in  $K$ ). That symmetry makes sense anyway, since (given that cost model) actual circuit complexity has the same symmetry. We’ll discuss irreversible 0-cost functions later.)

There is an interesting unconventional circuit-cost model (for general boolean computation) which considers a larger set of functions to have 0 cost — namely, all  $\mathbb{F}_2$ -affine functions — which results in the maximum possible complexity being roughly the square root of its usual value [BPP00]. This may have technical advantages in creating a useful pattern basis.<sup>15</sup>

---

<sup>15</sup> Treating all  $\mathbb{F}_2$ -affine functions (i.e. all XOR and NOT gates) as “free” suggests looking for a pattern basis which is symmetric not only for the  $n!$  wire permutations and  $2^n$  wire-negations (in

## 2.12 ... which has not yet been achieved...

So we know we want a pattern basis which makes all pattern maps have fairly sparse matrices, and simple ones sparsest, with the generating circuits' matrices almost as sparse as the identity matrix, but not quite (except for the 0-cost generating circuits). But we don't know any systematic way to construct a basis like that — if we did, we'd immediately get our useful complexity formula, assuming we understood the basis well enough to prove, for at least some functions of interest, that they matched enough of its patterns.

It's worth noting that if we could prove a useful pattern basis *existed*, in the sense of all rows of generating circuits' pattern matrices being sparse, then even if we only “understood” its “simple” patterns (treating most of the rest as beyond our specific understanding), we could potentially use it to prove significant lower bounds, by finding enough matching of simple patterns to truth tables of interest. We'd have to understand only enough about the non-simple patterns to prove the generating circuits didn't map any pattern's *dual* (i.e. the corresponding formal linear combination of truth tables, in the dual pattern basis, mapped from circuit output to circuit input) to a sum of too many other patterns' duals. (Such a sum corresponds to one row of the pattern matrix, so we're still talking about minimizing the number of nonzero values on each row.)

It's easy to keep this measurement low for single gates (such as S-boxes), simply because they have only a constant number of inputs; the hard part is to avoid increasing it (the number of nonzeros on any row) too much, as those circuits are adjoined to more and more non-interacting parallel wires, to construct generating circuits with more inputs and outputs. (We sometimes call this the *adjoined-wire issue*.)

(Most of what we've said so far should apply to most kinds of circuit models; but starting here, we'll be talking mainly about models intended for general boolean computation. Other models might be analogous in many ways, but the effect on the property space of “adjoining a non-interacting wire” (and even more, the considerations about how that relates to a useful pattern basis) would often be different than they are in the general boolean case we discuss below.)

---

the input space  $\mathbb{F}_2^n$ , whose coordinates correspond to “wires”), but for arbitrary  $\mathbb{F}_2$ -linear changes of basis of the input space (of which there are roughly  $2^{n^2}$ ). It also seems natural in conjunction with the “fourier matrix model of boolean computation” we'll discuss in Sections 2.16.3 and 5.2, in which any  $f: \mathbb{F}_2^n \rightarrow \mathbb{F}_2^m$  is viewed as an  $\mathbb{R}$ -linear map from XORs of inputs to XORs of outputs (but with each XOR-value  $b \in \mathbb{F}_2$  represented by  $(-1)^b \in \mathbb{R}$ ), which is representable by a  $2^m$  by  $2^n$  matrix of real-valued fourier coefficients of  $f$ ; these matrices multiply when functions are composed.

## 2.13 ... since naively “adjoining a wire” doubles measured complexity ...

The simplest way to try constructing a pattern basis is by induction over the number of wires  $n$ . That is, given a basis  $P_n$  for the property space  $Q_n$  over  $n$  wires (of dimension  $2^{2^n}$ ), we somehow construct a basis  $P_{n+1}$  for the property space  $Q_{n+1}$  over  $n + 1$  wires (whose dimension is  $2^{2^{n+1}} = (2^{2^n})^2$ ).<sup>16</sup>

The most straightforward version of that construction can be described (at the level of a given circuit’s pattern map, between its input and output property spaces) as simply ***tensoring the circuit’s pattern matrix with itself***, to get the pattern matrix for the same circuit “widened by one wire” — that is, with one new non-interacting wire adjoined in parallel.<sup>17</sup> (If we do this at every inductive level, we can end up with the indicator or monomial basis in each size of property space, depending on our choice of basis for the base case of 0-input truth tables.)

The problem with this construction is that as each wire is adjoined, the number of nonzero terms in each pattern matrix row is squared, which doubles the measured complexity. (If we generalized this framework to let the new wire have  $r$  possible values instead of 2, this construction would produce the  $r$ th tensor power of the pattern matrix, and thus multiply its measured complexity by  $r$  as each new wire was adjoined.)

If the new input wire had *not* been non-interacting, but (in addition to controlling the new output wire) also affected the rest of the circuit — effectively making it imitate one of 2 (or  $r$ ) unrelated circuits of the same complexity (with the new input value choosing which one) — then this effect of multiplying the circuit’s measured complexity by 2 (or  $r$ ) would be exactly right. But it’s a fatal flaw when the new wire is supposed to be non-interacting. In effect, this naive construction is treating every adjoined wire as if it affected everything in the circuit, resulting in a new circuit which is twice as complex (since its truth table is twice as large, and “presumed arbitrary”) — even when the new wire is actually non-interacting.

---

<sup>16</sup> It would be more precise to speak of  $n$  inputs rather than  $n$  wires, and (below) of the measured complexity of functions rather than of circuits — but “adjoining a wire” seems easier to visualize in the context of circuits. By the “measured complexity of a circuit”, we just mean the measured complexity of the function it computes, whether or not the given circuit is minimal for that function.

<sup>17</sup> If we define each  $P_{n+1}$  as  $P_n \otimes P_n$ , where  $P_n$  is the pattern basis set for properties of  $n$ -input 1-output truth tables, this also makes pattern matrices tensor with themselves, when one non-interacting wire is adjoined. In symbols (where function  $f'$  is  $f$  with one “wire” (new input and output) adjoined), we have (in this case)  $\text{PM}_{\mathcal{P}}(f') = \text{PM}_{\mathcal{P}}(f) \otimes \text{PM}_{\mathcal{P}}(f)$ . (See Section 2.17.4 for what we mean by “tensoring basis sets” like  $P_n$ , and Sections 3 and 4 for other notation and definitions used here.)

To fix this problem, the inductive construction has to be done in a more sophisticated way, which recombines or “mixes” (by an appropriate change of basis) the “naively”-tensored smaller-basis elements,<sup>18</sup> in a way which, when the pattern matrices being tensored are the same, doesn’t increase (too much, or preferably at all) the number of pattern matrix terms (per matrix row) needed to express any given generating circuit.

Finding a single change of basis (for each value of  $n$  in this inductive step), which would fix this problem for every generating circuit at once, would be hopeless, if it needed to work for arbitrary matrices. But the pattern matrices and pattern basis are far from arbitrary, and it seems plausible they can be given special properties which make this possible. There are at least two points we can take advantage of:

- As we’ll discuss near the end of Section 2.16.2 (about kinds of patterns), the pattern basis might be able to have high symmetry, even though most of its elements are “complex” — since *compared to arbitrary properties*, they can have *very low* complexity.
- Each 1-output function (truth table) of  $f$ ’s input or output space, when it’s expressed (uniquely) as a sum of dual patterns, can *also* potentially have high symmetry (e.g. with respect to permuting its coefficients in certain ways, along with multiplying them by roots of unity in  $K$ ), since there are only as many functions as dual patterns (i.e. as dimensions of the property space), so by no means do arbitrary combinations of coefficients need to be allowed. (Even considering the larger number of functions *between* the input and output spaces, their number is more like the number of dual patterns than like its exponential, so this point remains largely true.)

## 2.14 ... but there’s hope, since that’s the *only* hard issue we need to solve.

In a conventional circuit model for general boolean computation, it turns out we only need the pattern basis to satisfy three criteria, to be able to prove nontrivial complexity bounds for most functions (though this won’t guarantee we can do that for specific functions of interest):

1. it’s symmetrical for permutations of wires, so it sees them as having 0 complexity (or perhaps, “almost symmetrical”, so it measures their complexity as

---

<sup>18</sup> I.e.,  $P_{n+1} = W_{n+1}(P_n \otimes P_n)$  (where  $W_{n+1}$  is a square matrix we design, for each  $n$ ). If  $f'$  is  $f$  with one wire adjoined, their pattern matrices obey  $\text{PM}_{\mathcal{P}}(f') = W_{n+1}(\text{PM}_{\mathcal{P}}(f) \otimes \text{PM}_{\mathcal{P}}(f)) W_{n+1}^{-1}$ .

“small” — note that the term “complexity” as used here just means “whatever value our formula measures”, whether or not that relates to actual circuit complexity);

2. it measures most functions as having significant complexity (like a random basis would);
3. it handles the adjoined-wire issue (so adjoining wires leaves measured complexity unchanged, or at least doesn’t change it “too much”, as detailed below).

A conventional circuit model’s “free” operations include not only permuting wires, but discarding or splitting them, or creating new (constant-valued) wires. Ideally our pattern basis will agree (by giving those *trivial wire operations* a measured complexity of 0); but if not (as long as it meets the listed conditions), we can work around this by artificially treating them (in a modified circuit model) as having a small constant cost, which gives only a small linear weakening of our lower bounds.<sup>19</sup> However, I would guess that making those operations have 0 measured complexity in the first place will be easier than solving the adjoined-wire issue, and may help show how to solve it (since both issues involve the relationship between the pattern basis and the subalgebras of the property space consisting of properties definable on various subsets of wires).

The reason we consider “adjoined wires” the main issue is that solving the other two together doesn’t look hard. For each input or output space of size  $n$  (namely,  $\mathbb{F}_2^n$ , whose  $n$  coordinates correspond to “wires”), the wires have only  $n!$  permutations, so constraining an otherwise-arbitrary basis of the associated property space (of dimension  $2^{2^n}$ ) to be symmetrical under their action should still leave it a lot of freedom, for example to be “mostly random”.

---

<sup>19</sup> Technically this weakening is not precisely “linear”, since a function with arbitrary  $n$  and  $m$  which can be computed using *only* trivial wire operations (e.g. discard all inputs, then output all zero bits) would have 0 complexity in a conventional model, but arbitrarily high complexity if trivial wire operations were given constant positive cost.

But if we ignore or work around this issue (for example by first redefining conventional complexity to be higher by  $n + m$ ), then “linear weakening” is accurate: a lower bound of  $X$  on our revised complexity (with trivial wire operations of constant positive cost) implies a lower bound of  $\Theta(X)$  on the originally-defined (but with extra  $n + m$ ) circuit complexity (and the suppressed constant factors are small). Proof: a minimal circuit in the original cost, using  $X$  gates, needs at most  $n + m + cX$  trivial wire operations, for some small  $c$  depending only on the circuit model (namely, the maximum of  $n_g + m_g$  over the primitive gates, where gate  $g$  has  $n_g$  inputs and  $m_g$  outputs). The scaling factor change is also constant and small. The result follows. (A more formal but much longer version of exactly the same proof can be found in Section 4.7.1.)

(Of course, solving the adjoined-wire issue would presumably force the pattern basis to be highly *nonrandom*; but the fact that *most* bases would give most functions high measured complexity gives reason to believe that that feature (number 2 in the list above) is not impossibly constraining either.)

## 2.15 Proof outline: a pattern basis meeting *these* conditions will be useful.

Here is a more precise version of the statement about the “three criteria” made above, and a preview of the proof (which is given more formally in Section 4.6):

Suppose we construct a pattern basis which fixes the adjoined-wire issue in the ideal way — so adjoined (parallel, non-interacting) wires don’t affect measured complexity at all, at least for generating circuits — and for which permuting wires has 0 measured complexity (due to symmetry).

In a conventional circuit model for general boolean computation, we can get by with just one kind of generating circuit (aside from wire permutations): one small circuit (which might be either a primitive gate or a trivial wire operation) with zero or more adjoined wires.

The small circuits have low measured complexity (since their maximum size is a small constant of the circuit model), so by our assumption about the effect of adjoined wires, the generating circuits do too — their measured complexity is limited by a small constant. (This also solves the “arbitrary width issue” (about scaling measured complexity to fit the costs of the infinite set of generating circuits of arbitrary widths), as promised a few subsections back, since it means we only have to take into account the finite set of small circuits that form their “cores”.)

Since all generating circuits have small measured complexity, a small positive constant  $\lambda$  (dependent only on the circuit model and pattern basis) will scale that to fit their actual cost — and after scaling, measured complexity remains “subadditive” for function composition, so it becomes a valid lower bound on circuit complexity for all constructible functions in the model. Then for *every* function  $f$  with high measured complexity in this pattern basis, we immediately get a proportionally-high lower bound on  $f$ ’s actual circuit complexity.

(So, informally, a pattern basis meeting the above conditions will be “useful” if we can prove it measures high enough complexity on enough functions  $f$ , or enough “interesting” functions  $f$ , to consider that an interesting result. The above conditions guarantee its measured complexity for *all* constructible functions  $f$  is (up to a small constant factor) a lower bound on  $f$ ’s actual circuit complexity.) ■



(Even if we can't solve the adjoined-wire issue in the ideal way — that is, if adjoining a wire adds *some* measured complexity — we're ok for many purposes as long as the amount added can just be a fixed polynomial in  $n$ , so adding it needn't double the measured complexity of the original circuit. Similar comments might apply even if permuting wires is not free. We'll further discuss these “fallback” situations in Section 4.7.)

As a possible guide for finding a useful pattern basis, I'll describe in more depth the nature of the patterns we hope to get by choosing the pattern basis “correctly”, and how they might relate to the kinds of patterns we understand intuitively. (In later sections I'll describe the considerations which motivated this framework, and give more reasons we might expect finding a useful pattern basis to be possible.)

## 2.16 desired qualitative nature of patterns

This subsection summarizes the general picture I hope can be found, of the kinds of patterns that might exist in a useful pattern basis (for general boolean computation), how various kinds of functions would match them, and how enough provable matching might be found to lower-bound the complexity of interesting functions. (Later sections go into more detail about some of these points.)

### 2.16.1 kinds of functions

- In general, a function  $f$  of measured complexity  $X$  matches about  $c^X$  patterns for some smallish  $c$  (which I guess is constant, but we could survive if it was linear or even polynomial in  $n$ ). By  $f$  ***matches*** (or ***has***) pattern  $p$ , we mean the associated pattern value is either nonzero or non-“small”, depending on the matrix measure used to construct the complexity measure. (Nonzero seems more likely than non-“small”, due to the quantized nature of the truth table and therefore of the possible pattern values (especially if patterns have high symmetry as properties, as seems likely); and if  $K$  is a finite field rather than  $\mathbb{R}$  or  $\mathbb{C}$ , it's not obvious what “small” could mean (though something involving subfield membership might be possible). I think this means “nonzero” should be hoped for here, but it doesn't seem safe to completely assume.)
- The patterns differ in their complexity (for a meaning of “complexity” which is not fully precise here, but for which Kolmogorov complexity of their expression as a polynomial in the truth table entries (noting that the entry indices have

structure which matters here) can stand in for now, since we essentially only use it to avoid having to make explicit counting arguments); we say more about the nature of individual patterns below. An important point mentioned earlier is that any algorithm to probe an arbitrary truth table at a polynomial number of possible patterns (relative to the size of the truth table, about  $2^n$ ) is unable to scan non-simple patterns with any significant “density” (the fraction which get scanned).

- Low-complexity functions with special structure might have a predictable set of simple patterns. In some cases this lets us prove a lower bound on their complexity, by identifying enough patterns we can prove they have.
- High-complexity functions with special structure might or might not have a predictable set of complex patterns (as well as some simple ones). Again, if we can prove they have enough patterns, we can prove a nontrivial bound on their complexity.
- Pseudorandom functions are low complexity and thus have relatively few patterns, but (essentially by definition of “pseudorandom”) those patterns are all somewhat complex (though not maximally complex), and are predictable only if you know exactly how the functions are constructed (which, again by definition of “pseudorandom” [GGM86], must include complete knowledge of their construction parameters, including “random seed”, not just of the general scheme). Thus it’s hard to notice their relative lack of complexity (or anything specific about what kind of complexity they have) by probing either simple or randomly chosen patterns. (You could also just “look at the truth table and see if you notice anything”, but, we believe, this effectively probes only some simple patterns, not any complex ones.) (Presumably there is some theorem that lets you infer, solely from some function  $f$ ’s complete lack of simple patterns, that it must have at least some minimal complexity (probably roughly linear in  $n$ ), even though you can’t find any of the many patterns that theorem proves it must have.)
- “Random” functions (those of near-maximal circuit complexity, as is true of most functions) have *more* patterns than any other kind of function (and almost all of their patterns are very complex), but still there are far more patterns they *don’t* have (i.e., as with all functions, most of their pattern match values are zero or “small”).

(A few patterns might be not only “simple” but *non-discriminating*, i.e. they might match a substantial fraction of all functions, so they would be an exception to the general but approximate statement that a random or pseudorandom function matches *no* simple patterns.)

### 2.16.2 kinds of patterns

- Some of the patterns in a useful pattern basis would be simple ones we might recognize. They might include (or in some sense “correlate to”) the kinds of facts about a truth table for  $f$  which often have “visible consequences” when  $f$  is composed with some generating circuit — that is, which often lead to other patterns being visible in the truth table of the composition — such as  $f$ ’s “*density*”, or its correlation with specific very simple truth tables, or (more generally) the values of low-degree highly-structured polynomials in its entries.

For this purpose it’s often convenient to treat truth table entries as elements of  $\{1, -1\}$ , i.e. as  $(-1)^b$  for  $b \in \mathbb{F}_2$ ; we then have another basis of truth table properties consisting of monomials in those entries, each of which is the XOR (product, in this representation, corresponding to the sum in  $\mathbb{F}_2$ ) of some fixed subset of entries. Using this representation, the density of a truth table is just the average value of all entries, and its correlation with any fixed truth table is a degree-1 polynomial in the entries; the  $2^d$ th power of the  $d$ th Gowers uniformity norm<sup>20</sup>, and the “influence” of any set of  $d$  inputs [KKL88], are highly-structured degree- $2^d$  polynomials. (A caveat: these examples, and some others in this subsection, assume the property space is based on a field  $K$  in which density and correlation values can be expressed directly, i.e.  $\mathbb{R}$  or  $\mathbb{C}$ . If  $K$  was a finite field (and especially if its characteristic was 2), other examples would be more suitable, though some of them taken in combination could still correlate to the kind of property normally expressed with values in  $\mathbb{R}$ .)

- An interesting class of simple patterns are the  $2^{m+n}$  *fourier coefficients* [Odo14], each defined as the expected product (under the  $(-1)^b$ -representation) of  $f$ ’s output (or the XOR of some subset of its outputs) and the XOR of some subset of its inputs. These can be organized as a *fourier matrix* which represents  $f$  as a linear map from  $\mathbb{R}^{2^n}$  to  $\mathbb{R}^{2^m}$  (which we’ll discuss later as the “fourier matrix model”). This involves only tiny subspaces of the corresponding property spaces (of dimension  $2^n$  rather than  $2^{2^n}$ ), but it’s worth pointing out

---

<sup>20</sup> That is,  $\|\cdot\|_{U^d(\mathbb{F}_2^n)}^{2^d}$ ; this was discussed as a property of boolean functions in [Gow09]. The norm  $\|\cdot\|_{U^d}$  was introduced in [Gow01]; see [WikG] for notation.

that within them, we can already do a “miniature version” of what we want to do in the property space as a whole. Specifically, the fourier matrix of a  $w$ -input gate (plus any number of adjoined wires) has at most  $2^w$  nonzero entries per row (out of  $2^n$ ); as we compose circuits that number multiplies, so it grows at most exponentially (a similar point was made in [Gow09a]). This lets us prove (trivial) circuit size lower bounds of the form  $\log(2^q/2^w) = \Omega(q)$ , where  $q$  depends on  $f$  but can be no more than  $n$ . If we could extend this behavior to the whole property space, those bounds could instead be more like  $\log(2^{2^q}/2^{2^w}) = \Omega(2^q)$ , where  $2^q$  can be no more than  $2^n$ . So in a sense, all we need to do is choose all the higher-degree patterns so they share this important feature of the (degree-1) fourier coefficients.<sup>21</sup>

- Other properties can be made by combining those patterns (using sum and product, i.e. forming polynomials of them), which (if those polynomials are either small, or highly structured) might exist directly in the pattern basis, or might be relatively small linear combinations of patterns in it. Properties like this can express things like “ $f$  approximates a certain function of a few of its inputs” (or “of a few XORs of subsets of its inputs”) — the kind of pattern whose presence in  $f$ ’s truth table intuitively implies the need for any circuit computing  $f$  to somehow do “a certain piece of computational work”.
- It’s even possible to approximate *any function* of simpler patterns (not just a polynomial of them) using relatively few terms in linear combination, by means of a short Taylor series. It’s also potentially possible for any expressible property to be included in the pattern basis directly — for example, a property like “log of the fraction of all truth tables which have the same density as this one” — provided not too many are included, and they’re linearly independent in the property space. If there is some reason for properties like that to be in the pattern basis, or to be expressible as small linear combinations of its elements, we might be able to choose it to make that true. (For reasons like these, the pattern basis seems likely to be either closed under product of patterns (where that product is defined as the pointwise product of pattern values, treating patterns as functions of truth tables — note that this is not the same as *composition* of patterns (when the functions containing them are composed), though in other contexts we might also treat that as a product), or “approximately closed” in the sense that products have relatively small expressions as linear combinations of pattern basis elements.)

---

<sup>21</sup> This doesn’t mean the fourier coefficients themselves must be elements of a useful pattern basis, but it seems likely some patterns would be closely related to them.

- But most patterns (not only in the pattern basis, but in *any* basis of a property space) would be very complex; note that there are  $2^{2^n}$  patterns, so most of them have a Kolmogorov complexity of nearly  $2^n$  bits — as much as a random truth table.

An important observation related to this: Any pattern  $p$  which happens to match some random function  $f$  (where  $f$  is random in the sense of having near-maximal Kolmogorov complexity), but which is non-matching on almost all functions, *needs* a high Kolmogorov complexity itself, or the very fact of its matching will force  $f$  to be non-random, since otherwise we could “compress”  $f$  by describing it as “the  $k$ th function which matches  $p$ ”. Or more precisely, when we *do* describe  $f$  that way, it must be true that  $C_K(k) + C_K(p) + C_{K0} \geq C_K(f)$ , where  $C_K()$  denotes Kolmogorov complexity and  $C_{K0}$  is a small constant; but  $C_K(k)$  can’t be too large a fraction of  $C_K(f)$  if  $p$  is non-matching for almost all  $f$  (since then  $k$  can’t be nearly as large as the number of possible functions  $f$ ). Thus the only way to realize the vision of random functions being matched by *any* highly discriminating pattern, not to mention by *lots* of such patterns, is to have lots of high-complexity patterns.

- On the other hand, relative to an *arbitrary* property of an  $n$ -bit truth table (of which there are  $k^{2^{2^n}}$  whose values are in any given  $k$ -element subset of  $K$ ), the elements of the pattern basis can all be “exceptionally simple”.

The last point means we have plenty of “room” to be very selective about which properties we include in the pattern basis — for example, it may well be able to be a highly symmetrical structure made only of special kinds of properties, in spite of most of them being as complex as a random truth table. Indeed, as mentioned earlier, we want it to include only those properties (of a function  $f$ ) which are the most significant for predicting *other* patterns in compositions of  $f$  with generating circuits — but which are “significant in independent ways”, meaning not only that the patterns are linearly independent in the property space, but that relatively few pairs of them interact directly, in the sense of having nonzero entries in the pattern matrices of simple functions.

That criterion for a good pattern basis is both strong and self-referential (though it’s not vague, when expressed as pattern matrix sparseness for all simple boolean functions); but the relative hugeness of the property space as a whole would seem to give some hope that it can be met.

### 2.16.3 lower-bounding the complexity of specific functions

Assuming a good pattern basis could be found, how might we prove a reasonable lower bound on measured complexity for a function with special structure?

To be more specific: given a well-understood function family  $f = \langle f_n \rangle$ , where  $f_n: \mathbb{F}_2^n \rightarrow \mathbb{F}_2^m$  has “intuitive complexity”  $X$  (and  $m$  and  $X$  depend on  $n$ ), how might we prove  $f_n$  has on the order of  $c^X$  patterns (for some small constant  $c$ )? (For this discussion, it’s sufficient to assume  $f$  has just one output (i.e.  $m = 1$ ). It’s also worth noting that even if we find only  $c^{\sqrt{X}}$  or  $c^{X^\alpha}$  patterns (for some  $\alpha > 0$ ) — due either to limitations in our pattern basis, or to  $f$  being less complex than we think — we could still get significant bounds like  $\text{NP} \not\subseteq \text{P/poly}$ , provided  $X$  is superpolynomial in  $n$  and  $f$  is in  $\text{NP}$ .)

As an example, we’ll let  $f$  be a “**feature detector**”, whose output is true whenever its input matches any of  $X$  seemingly independent “features”  $F_i$  (predicates over  $f$ ’s input space  $\mathbb{F}_2^n$ ).<sup>22</sup> Since we want  $X > n$ , not all combinations of input features can occur — they can’t be *jointly* independent. Even so, by making them *pairwise* independent (and giving them other properties we’ll define below), we’ll be able to make an intuitive argument for  $f$  having high complexity.

Though we can’t formalize this argument (let alone prove it), it will suggest a way in which  $f$  might be able to match enough patterns in a useful pattern basis. Specifically, we’ll argue that the following steps are plausible: for each input feature  $F_i$ , find a set of  $c$  properties  $P_i$ , each of which  $f$  matches due to its “approximately detecting” feature  $F_i$ ; observe that  $f$  matches all  $c^X$  *products* of those properties (since the product of their (nonzero) values at  $f$  is nonzero), and argue that many of those product properties are distinct; finally (using our guesses about the nature of a useful pattern basis), derive from these a sufficient number of related *patterns*, also matched by  $f$ .

(Note that we’re not suggesting this is the *only* way to prove an explicit function matches lots of patterns, nor are we claiming it can surely be made to work. We’re just arguing that a conclusion like this might be plausible, by speculating on one possible way of reaching it.)

When thinking about possible patterns, or reasoning (intuitively) about complexity, we’ll consider **input-XORs** (XORs of subsets of  $f$ ’s inputs) to be just as fundamental as single inputs. In other words, we won’t pick a preferred basis in  $f$ ’s

---

<sup>22</sup> As we’ll see after defining  $\langle F_i \rangle$ ,  $f$ ’s *true* complexity will turn out to be at most  $\tilde{O}(\sqrt{X})$ . (Later we’ll mention a possible partial explanation for this discrepancy.) This needn’t discourage us from trying to prove a superpolynomial lower bound, since  $\sqrt{X}$  is superpolynomial whenever  $X$  is.

input space  $\mathbb{F}_2^n$  (considered as a vector space over  $\mathbb{F}_2$ ). This will help us define a superpolynomial number of input features, since (over input vectors in  $\mathbb{F}_2^n$  chosen uniformly at random) all pairs of the  $2^n$  input-XORs are statistically independent.

Our main justification for this viewpoint is the ***fourier matrix model of boolean computation*** (which we’ll also discuss in Section 5.2; this should not be confused with the exponentially-larger “***property space model***”, which is a major subject of this paper<sup>23</sup>). In the fourier matrix model, we represent all bits  $b \in \mathbb{F}_2$  (whether they’re inputs, outputs, or boolean function values) by  $(-1)^b \in \mathbb{R}$ , so XOR is multiplication in  $\mathbb{R}$ , and the correlation between two ***signals*** (arbitrary boolean functions over  $\mathbb{F}_2^n$ ) is their expected product (over random input vectors). This lets us represent any boolean function or circuit (from  $n$  inputs to  $m$  outputs) by a linear map (from  $\mathbb{R}^{2^n}$  to  $\mathbb{R}^{2^m}$ ) between its input-XORs and output-XORs (or equivalently, by a matrix of all its fourier coefficients; composing functions composes the linear maps and multiplies their matrices).<sup>24</sup>

This model lets us see any circuit  $C$  which computes  $f$  as some way of *linearly* moving and combining  $C$ ’s input-XORs, placing them in turn onto the wire-XORs of  $C$ ’s intermediate wires at each “stage” (single gate, and all wires alongside it), so as to end up with just the desired combination of input-XORs on  $C$ ’s output — namely, the fourier components of  $f$ . To the extent that  $f$  approximates some function  $f'$  of the input-XORs in some subspace  $W$  of  $(\mathbb{F}_2^n)^*$  (the dual of  $f$ ’s input space, as a vector space over  $\mathbb{F}_2$  — this is just the space of all the input-XORs of  $f$ , or of  $C$ ), we can argue (intuitively), after noting that all fourier components of  $f'$  are (represented by) elements of  $W$ , that the only “computational work” (linear motions/combinations of signals within the circuit) which is “useful” for  $f$ ’s approximation of  $f'$  is the part involving the input-XORs in  $W$  (as potential fourier components of any wire-XOR in the circuit — that is, if we were to formalize “this part of the work”, the first step would be to project every signal in the circuit into the part containing only those

---

<sup>23</sup> Though we won’t use this “fourier matrix model” except to support this section’s intuitions about feature-detector functions, it’s worth pointing out that it’s *related* to the property space model, in the sense of being a “submodel” or “projection”, obtainable by ignoring all properties except degree-1 (homogeneous) polynomials in truth table entries (represented as  $(-1)^b$ ) (provided we’re using property values in  $\mathbb{R}$  or  $\mathbb{C}$ ). This raises the possibility that the vague ways we’ll use it here might be more successfully formalizable if we could extend them into the larger model.

<sup>24</sup> [Odo14] also discusses using this  $(-1)^b$ -representation to compute fourier coefficients, XOR, and correlation of 1-output boolean functions, and the fact that the  $2^n$  parity functions (what we call input-XORs) form an orthonormal basis of (what we call) the “vector space of signals over  $n$  inputs”,  $\mathbb{R}^{2^n}$ . But I’m not aware of any prior publication of the fourier matrix model itself, or of any representation of boolean function composition by matrix multiplication, except for the special case of reversible functions, which can also be considered a special case of quantum computation [NC00] (an example of this correspondence, involving fourier coefficients, was noted in [Gow09a]).

fourier components).

Though we can't fully formalize “computational work” or our reasoning about it, this viewpoint suggests intuitively that if some circuit  $C$  simultaneously approximates two functions  $f'$  and  $f''$  (in a way we'll describe more precisely below), each determined by the input-XORs in independent  $\mathbb{F}_2$ -subspaces  $W'$  and  $W''$  of  $(\mathbb{F}_2^n)^*$ , then there is a limit to how much the “work” needed for  $C$  to approximate each of  $f'$  and  $f''$  can “overlap” (make use of the same ( $\mathbb{F}_2$ -subspaces of) wire-XORs in  $C$ ). This is because  $f'$  and  $f''$  are statistically independent (over random input vectors of  $C$ ), and so are any other functions of the input-XORs in  $W'$  and  $W''$  respectively; but those are the only intermediate functions we recognized as “useful” for their respective computations (getting  $C$  to approximate  $f'$  and  $f''$ ), and which we therefore defined as part of that “work”. But any one wire-XOR's ability to be correlated with independent signals (such as all nontrivial input-XORs in  $W'$  or  $W''$  respectively) is limited, since the sum of the squares of its fourier coefficients is limited to 1.<sup>25</sup>

The above informal reasoning is far from conclusive; as far as I know, there is no easier way to usefully formalize it than to somehow extend it into the exponentially-larger property-space model. But it does suggest the following construction of our example function. After spelling that out, we'll slightly improve our justification of its “intuitive complexity”, and then discuss it in relation to a hypothetical pattern basis.

To construct specific feature predicates  $F_i$  for our example function family  $f = \langle f_n \rangle$ , we'll use parameters  $w$ ,  $k$ , and  $X$  to be chosen later (as functions of  $f$ 's input size  $n$ ). Abstractly, we'll follow these steps:

1. Choose  $X$  pairwise-independent  $w$ -dimensional subspaces of  $(\mathbb{F}_2^n)^*$ . We'll call the subspaces  $W'_i$ ; in each one, choose a basis  $W_i$ , which will be a  $w$ -element set of input-XORs of  $f$ .
2. Define each  $F_i$  as some fixed function of the (values of the)  $W_i$ , which is true (i.e., the  $i$ th input feature is present) for only one combination of values (so for random input vectors, it will be true  $1/2^w$  of the time). Choose parameters so that for most input vectors, every  $F_i$  is false (and thus  $f$  is false).
3. To ensure  $f$  is in NP, it's enough to do all this using a systematic structure, so we can nondeterministically guess an “index”  $i$ , and efficiently check whether the  $F_i$  it indexes is true.

---

<sup>25</sup> The fact that this limit is on a sum of *squares*, rather than of *absolute values*, is one possible source of the discrepancy mentioned earlier between our example function's “intuitive complexity” of  $X$  and the known upper bound of  $\tilde{O}(\sqrt{X})$  on its actual complexity.



To do this concretely, organize the  $n$  inputs of  $f$  into a matrix with  $w$  rows (ignoring “leftover” inputs if  $n$  is not a multiple of  $w$ ); then define  $f$  as true whenever some subset of  $k$  columns (each treated as a “column vector” in  $\mathbb{F}_2^w$ ) XORs to the 0 vector (where XOR means vector addition in  $\mathbb{F}_2^w$ ).<sup>26</sup> Or in the terms we used above: each index  $i$  picks some subset of  $k$  columns in the matrix of inputs; the associated set  $W_i$  contains, for each of the  $w$  rows of that matrix, the XOR of the  $k$  inputs in that row and the  $k$  chosen columns; each  $F_i$  is true whenever all  $w$  of those XORs (i.e., all elements of the associated set  $W_i$ ) are 0.

The number of input features  $F_i$  in this construction is  $X = \binom{\lfloor n/w \rfloor}{k}$ . If  $k$  slowly increases with  $n$  and  $n \gg w \gg \log_2 X$  (for example,  $k = \log n$  and  $w = k^3$ , rounding up to integers as needed), then  $X$  is superpolynomial in  $n$ , but for most input vectors every  $F_i$  is false. (As alluded to earlier, there are circuits of size  $\tilde{O}(\sqrt{X})$  which can evaluate this function, by sorting the XORs of each subset of  $k/2$  column vectors, then looking for repeated values (which will be adjacent in the sorted list) that don’t involve overlapping column subsets.)

We can now make more precise (though it will remain informal) the sense in which  $f$  “approximately detects” each input feature  $F_i$ , and therefore (intuitively) seems to require its circuits to do “most of the computational work” needed to fully detect that feature (and furthermore, that this “work” is pairwise “independent” for  $i \neq j$ ).

Imagine repeatedly running  $f$  on random input vectors chosen uniformly from  $\mathbb{F}_2^n$ , and asking how close it comes to computing a given  $F_i$ . Since  $F_i$  depends only on the values of the  $w$  XORs in  $W_i$ , we’ll call them “important input signals”, and call everything else about  $f$ ’s input vector “random noise”. We might then ask: “conditioned on the values of the XORs in  $W_i$  (with the input to  $f$  otherwise random), how good is  $f$  as a measuring device for  $F_i$ ?” (Note that even if we extend our definition of “important” input signals, by “XOR-closure”, to cover the entire  $\mathbb{F}_2$ -subspace  $W_i'$ , so that  $F_i$ ’s “ $(-1)^b$ -representation” is an  $\mathbb{R}$ -linear combination of “important” signals’  $(-1)^b$ -representations, this won’t affect what we’re conditioning on, since the extended signals are functions of the original ones.)

The answer is,  $f$  is a *very* good measuring device for  $F_i$  — for *any* specified set of values of the  $w$  XORs in  $W_i$ , the probability that  $f$  and  $F_i$  disagree is bounded above by the (conditional) probability that any  $F_j$  (for  $j \neq i$ ) is true, which (by the union bound) is at most  $\sum_j 2^{-w} = (X - 1)/2^w \ll 1$ . (Note that the pairwise independence of the  $W_i'$  is crucial for this conclusion, since it tells us that conditioning on values of  $W_i$  has no effect on values of  $W_j$ .)

---

<sup>26</sup> This is reminiscent of the “d-SUM problem” (in which the column vectors would instead be added as signed  $w$ -bit integers), which is well-known (according to [PW10]).

That completes our intuitive argument for any circuit  $C$ , which computes  $f$ , necessarily “doing most of the work” of detecting each  $F_i$ . But the same argument also shows statistical independence between *any* function of the  $W_i$  and *any* function of the  $W_j$  (for any  $i \neq j$ ), which is just the intuitive argument we described earlier for the “work” associated with approximately detecting  $F_i$  being “independent” of the similar “work” associated with  $F_j$ .

Now we turn to the question of finding patterns in our example function  $f$ . For this, we’ll leave aside the vague concept of “computational work”, but replace it with the hypothetical (though not as vague) assumption that we’ve found a “useful” pattern basis with some of the properties we hope for, especially in connection with how it relates to the vector space structure of  $f$ ’s input space  $\mathbb{F}_2^n$ , and how it solves the “adjoined-wire issue”. In light of our discussion above, we’ll assume our pattern basis is symmetrical (up to scalar multiples) with respect to any change of basis in  $\mathbb{F}_2^n$  or  $\mathbb{F}_2^m$ , as well as any negation of inputs or outputs — in other words, with respect to composing  $f$  with any  $\mathbb{F}_2$ -affine function.<sup>27</sup>

Near the beginning of this subsection, we outlined steps we hope to show are plausible for finding on the order of  $c^X$  patterns in  $f$ . The first step is to find, for each input feature  $F_i$ , a small set of properties  $P_i$ , which  $f$  has due to its approximate detection of  $F_i$ . In light of our discussions above and in Section 2.16.2, it should be clear that for any property of  $F_i$  itself (considered as a function of the smaller input space  $\mathbb{F}_2^w$  corresponding to the values of the  $W_i$ ) — or more precisely, for a property of functions from  $\mathbb{F}_2^w$  to  $\mathbb{R}$  (which we can apply to the expected value of  $f$ , conditioned on the values of the  $W_i$ ) — there are many related properties of  $f$  (which measure things related to how well, in various senses,  $f$  approximates various functions of the  $W_i$ ); for example, various sums and products of (the properties which measure) the  $2^w$  fourier coefficients of  $f$  which are nonzero in  $F_i$ . These provide many candidates for elements of  $P_i$ .

Next, we observed that if  $f$  has several properties  $\langle p_i \rangle$ , it also has their pointwise product<sup>28</sup> property  $\prod_i p_i$ . (The reason is simply that all it means for  $f$  to “have” a property  $p$  is for  $p$ ’s value at  $f$  to be nonzero.) This already gives us  $c^X$  properties which  $f$  has (where  $c$  is the minimum size of any  $P_i$ ). (These product properties might not all be distinct, but to the extent that the values of all  $W_i$  taken together can distinguish many input vectors from one another, it seems likely that enough of them are distinct.)

---

<sup>27</sup> See also our discussion of [BPP00] and related topics, in Section 2.11 and its footnote.

<sup>28</sup> This “pointwise” product of properties is defined by treating them as functions of truth tables; it should not be confused with the product (of dual properties) coming from function composition.

By itself, this is not yet interesting — for any arbitrary property,  $f$  is likely to have it. What’s interesting is finding lots of *patterns*  $f$  has, not just lots of properties. What we need is some connection between the properties we’ve mentioned, and the hoped-for structure of a useful pattern basis.

Recalling our discussion of the adjoined-wire issue, we hoped to solve it, as part of the inductive construction of a pattern basis, by an appropriate change of basis in a property space which recombined patterns defined in its subspaces in a useful way. Combining this with our hoped-for symmetry of a pattern basis with respect to changes of basis in  $\mathbb{F}_2^n$ , something similar should hold for the relationship between patterns in the property spaces for  $\mathbb{F}_2^n$  and for the (dual) subspaces  $W'_i$  defined above. The hope is that, once a specific useful pattern basis is understood, its relationship to candidate properties for these sets  $P_i$ , and their product properties, will be simple enough to find enough patterns in  $f$  related to sums of those product properties.

Of course this hope (not to mention the evidence for its being achievable) is quite vague; but I think these arguments show that it’s at least plausible (that is, not obviously impossible) that an explicit function family similar to this example could provably have enough patterns to give us a nontrivial lower bound on its complexity.

## 2.17 Figures and examples

These subsections give examples of pattern matrices, and illustrate how they can be constructed and used. They can be skipped if desired, since most of the same information is described elsewhere, both informally (earlier in this Summary), and formally (Sections 3 and 4, which also define some notation we use here).

### 2.17.1 computing measured complexity ( $C_{M,\mathcal{P}}$ )

The table on the next page, read from the top down, gives an overview of computing the pattern matrix, and from it the measured complexity, for some tiny example functions, each of the general form  $f: \mathbb{F}_2^n \rightarrow \mathbb{F}_2^m$  (that is, a boolean function from  $n$  inputs in  $\mathbb{F}_2 = \{0, 1\}$  to  $m$  outputs in  $\mathbb{F}_2$ ).

Each function is described in the table by a circuit diagram, and equivalently by a tuple of output formulae; these are boolean formulae in its inputs  $x$  and  $y$ . (For example, the two-output function  $(\bar{x}, y)$  sets its first output to the logical negation of its first input  $x$ , and sets its second output to a copy of its second input  $y$ .)

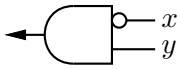
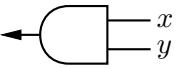
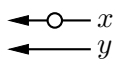
In this table, the functions all have exactly 2 inputs, and either 1 or 2 outputs. If these functions were part of a larger circuit, they would have to be “widened” by adjoining non-interacting parallel wires to account for all wires “passing alongside them” in that circuit; this would give them more inputs and outputs, and thus make their pattern matrices larger.

The function shown in the left column  $(\bar{x} \wedge y)$  is the composition of the two functions to its right — or as the table shows in symbols,  $(\bar{x} \wedge y) = (x \wedge y) \circ (\bar{x}, y)$  (with the last function,  $(\bar{x}, y)$ , having two outputs). This lets the table’s lower rows show the relationships implied by that composition, such as matrix multiplication equalities, or the subadditivity of measured complexity which results from that. (The table also shows “fourier matrices” (Section 2.16.3) for comparison, though they’re not used for computing measured complexity.)

We show the measured complexity  $C_{M,\mathcal{P}}(f)$ , defined as  $\log_2 M(\text{PM}_{\mathcal{P}}(f))$ , for one choice of pattern basis  $\mathcal{P}$ , and two choices of matrix measure  $M$ . As we’ve explained elsewhere, its value depends on  $M$  and  $\mathcal{P}$ , and (though always subadditive) is not necessarily related to actual circuit complexity. To become a lower bound on actual complexity, it must be scaled to fit generating circuit costs (of all allowed circuit widths); for this to be possible, and to result in nontrivial bounds, requires a carefully chosen pattern basis (whose existence has not yet been proven for  $m$  or  $n > 2$ ).

The pattern basis used here,  $\mathcal{P}_{\text{Rm2b}}$ , is only defined (so far) up to  $m, n \leq 2$ . (See Section 2.17.2 for more information about  $\mathcal{P}_{\text{Rm2b}}$ , and Section 2.17.5 for a comparison with other pattern bases, some of which can be defined for any  $m$  and  $n$ .)

**Table 1:** how measured complexity is computed (and why it's subadditive)

circuit diagram		=		o	
output formulae	$(\bar{x} \wedge y)$	=	$(x \wedge y)$	o	$(\bar{x}, y)$
$(m, n) = (\# \text{out-puts}, \# \text{inputs})$	$(1, 2)$		$(1, 2)$		$(2, 2)$
fourier matrix dims $(2^m \times 2^n)$	$2 \times 4$		$2 \times 4$		$4 \times 4$
fourier matrix (column order is $1, y, x, xy$ )	$\begin{vmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{vmatrix}$	=	$\begin{vmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{vmatrix}$	$\times$	$\begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{vmatrix}$
pattern matrix dims $(2^{2^m} \times 2^{2^n})$	$4 \times 16$		$4 \times 16$		$16 \times 16$
pattern matrix ( $\text{PM}_{\mathcal{P}}$ ), using a pattern basis $\mathcal{P} = \mathcal{P}_{\text{Rm2b}}$ described below, defined for $m, n \in \{0, 1, 2\}$	$\begin{vmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & -\frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & -\frac{1}{2} & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{vmatrix} =$		$\begin{vmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{vmatrix} \times$	$\begin{vmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{vmatrix}$	
computation of two variants of measured com- plexity $C_{M, \mathcal{P}}(f)$ :					
$M_{\text{nz}}$ of matrix	4	$\leq$	4	$\times$	1
$C_{M, \mathcal{P}} = \log_2 M_{\text{nz}}$	2	$\leq$	2	+	0
$M_{\text{abs}}$ of matrix	2	$\leq$	2	$\times$	1
$C_{M, \mathcal{P}} = \log_2 M_{\text{abs}}$	1	$\leq$	1	+	0

### 2.17.2 examples of pattern bases

Here are the change-of-basis matrices which define several pattern bases. (The details of this representation are described in a later subsection.) The first two are defined for all  $n$ ; the others are defined only for the values of  $n$  shown here.

- The indicator basis  $\mathcal{P}_{\text{ind}}$  and the monomial basis  $\mathcal{P}_{\text{mon}}$  were described earlier in the Summary, and are easy to define for all  $n$  (using identity matrices, or discrete fourier matrices over the group  $\mathbb{F}_2^{2^n}$ , respectively):

$$P_0^{(\text{ind})} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}, \quad P_1^{(\text{ind})} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix}, \quad \dots$$

$$P_0^{(\text{mon})} = \frac{1}{\sqrt{2}} \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix}, \quad P_1^{(\text{mon})} = \frac{1}{2} \begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{vmatrix}, \quad \dots$$

- The basis  $\mathcal{P}_{\text{Rm2b}}$  was discovered as part of a classification (to be published separately) of all “nice” real pattern bases defined for  $n \in \{0, 1, 2\}$ , where “nice” means orthonormal (in the canonical basis of each property space) and giving all  $\mathbb{F}_2$ -affine functions a measured complexity of zero. It was used to compute the pattern matrices shown in the table above.

$$P_0^{(\text{Rm2b})} = \frac{1}{\sqrt{2}} \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix}, \quad P_1^{(\text{Rm2b})} = \frac{1}{\sqrt{2}} \begin{vmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{vmatrix},$$

$$P_2^{(\text{Rm2b})} = \frac{1}{2\sqrt{2}} \begin{vmatrix} 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \\ 0 & -1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & -1 & 0 & 0 & 1 & 1 & 0 & 0 & -1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & -1 & -1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & -1 & 1 & 0 & 0 & -1 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & -1 & 0 & 0 & 1 & -1 & 0 & 0 & 1 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 & 0 & 1 & -1 & 0 & 0 & -1 & 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 & 1 & 0 & 0 & 1 & -1 & 0 & 0 & -1 & 0 & -1 & 1 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 \end{vmatrix}$$

(For an explanation of  $P_2^{(\text{Rm2b})}$ 's structure, and for more example pattern bases (some definable for all  $n$ ), see the separate classification mentioned above.)

- If we allow matrix entries in  $\mathbb{C}$ , there is another “nice” basis defined for  $n \in \{0, 1\}$ , which I call  $\mathcal{P}_{C1}$ . (It can be extended in at least one way to a nice but useless basis for all  $n$ , but I don’t yet know whether it can be extended usefully to  $n = 2$ . In this context, “useful” just means “nice, but giving nonzero measured complexity to AND”, since the “adjoined wire issue” can’t be studied for nontrivial functions until  $n = 3$ .)

$$P_0^{(C1)} = \frac{1}{\sqrt{2}} \begin{vmatrix} 1 & -i \\ 1 & i \end{vmatrix}, \quad P_1^{(C1)} = \frac{1}{\sqrt{2}} \begin{vmatrix} 1 & 0 & 0 & -i \\ 0 & 1 & -i & 0 \\ 0 & 1 & i & 0 \\ 1 & 0 & 0 & i \end{vmatrix}$$

- Finally, Table 2 (which shows measured complexities for various examples of pattern basis) includes the following randomly chosen real orthonormal pattern basis,<sup>29</sup>  $\mathcal{P}_{\text{rand1}}$ , mainly to show how comparatively badly it behaves, but also to demonstrate that the resulting measured complexity is still subadditive:

$$P_0^{(\text{rand1})} = \begin{vmatrix} 0.7423 & 0.6701 \\ 0.6701 & -0.7423 \end{vmatrix}, \quad P_1^{(\text{rand1})} = \begin{vmatrix} -0.1861 & 0.2500 & 0.9456 & 0.0932 \\ -0.7947 & -0.5938 & 0.0130 & -0.1255 \\ -0.1048 & 0.3376 & -0.0177 & -0.9353 \\ -0.5682 & 0.6863 & -0.3246 & 0.3175 \end{vmatrix}, \quad P_2^{(\text{rand1})} =$$

0.0401	0.1033	0.4851	0.0579	0.0534	-0.0545	0.2001	-0.2051	-0.2003	0.1588	0.1461	-0.2735	-0.0270	0.0970	-0.4995	0.4899
0.1799	0.1515	-0.0305	0.3457	0.2003	0.1240	-0.1366	-0.1682	0.0574	0.4958	0.1048	-0.2710	0.2377	-0.4962	-0.0040	-0.2926
-0.2397	0.2374	-0.1010	-0.2713	-0.1289	0.3596	0.5107	0.1346	0.1842	-0.0054	0.1301	-0.0524	0.5524	-0.0720	0.0479	0.1060
-0.2206	-0.2561	-0.1878	-0.0913	-0.0906	0.3089	0.1046	-0.7256	-0.0383	0.1245	0.1685	0.1292	-0.1349	0.2131	-0.1247	-0.2445
-0.3093	0.2237	0.0215	-0.0378	-0.3879	-0.1652	-0.2471	0.1571	0.3165	-0.0378	0.4255	0.1125	-0.1906	-0.2128	-0.4264	-0.1736
0.3320	-0.0218	-0.2168	-0.5501	0.1374	-0.3821	-0.2969	-0.2484	0.1021	0.0556	0.1572	0.0973	0.3706	-0.0093	-0.0909	0.1777
0.3285	0.0175	-0.1584	0.0547	0.2245	0.1865	0.0326	0.0194	0.3511	-0.3985	0.3868	-0.5010	-0.1901	0.2343	0.0110	-0.0572
-0.0340	-0.1974	-0.5539	0.2068	-0.3774	-0.3196	0.2262	-0.0111	-0.0194	0.2083	0.0616	-0.2733	-0.1458	-0.1100	0.1300	0.3768
-0.3937	0.1912	0.0712	0.1167	-0.0704	0.0816	-0.5090	-0.1852	0.3240	0.0877	-0.2200	-0.2549	0.1478	0.3067	0.2495	0.2839
-0.3651	-0.2046	0.3340	-0.2909	0.2958	-0.3435	0.2287	-0.1100	0.2126	0.0340	0.1068	-0.1966	-0.2108	-0.2773	0.3665	-0.0820
-0.1158	-0.3351	-0.1537	-0.1068	0.2202	-0.0429	0.0851	0.2866	0.3236	0.2979	-0.4000	-0.1700	0.0213	0.2113	-0.4900	-0.1787
-0.1534	-0.1713	0.0969	-0.0191	-0.1826	-0.2198	-0.1115	0.1779	-0.4626	0.0266	0.2690	-0.3673	0.3446	0.3388	0.0705	-0.3888
-0.0047	-0.6955	0.1873	0.1598	-0.1443	0.2070	-0.1961	0.0064	0.1120	-0.3116	0.0483	0.0035	0.2969	-0.3298	-0.0858	0.1927
0.4036	-0.1441	0.3183	-0.3010	-0.4539	0.2433	-0.0313	0.1434	0.1884	0.4376	0.0412	-0.0331	-0.1965	0.1119	0.2373	-0.0195
0.1226	-0.0723	0.1774	0.4682	0.0238	-0.3267	0.2514	-0.0513	0.3845	0.1070	0.2085	0.3956	0.2522	0.3448	0.1064	-0.0536
-0.2043	-0.1577	-0.1465	0.0018	0.4103	0.2512	-0.1856	0.3385	-0.1491	0.3292	0.4727	0.2439	-0.1115	0.1043	0.0945	0.2926

<sup>29</sup> When using  $\mathcal{P}_{\text{rand1}}$  in making Table 2, its change of basis matrices were “re-orthonormalized” (to undo the effect of rounding when printing at limited precision), by subtracting from each row some multiple of each prior row (to make the matrix orthogonal), then normalizing that row. In principle, this is not needed, since even a set of non-orthogonal change of basis matrices should induce a “valid” (that is, subadditive) measured complexity function.

### 2.17.3 pattern matrix for the AND function

As a detailed example of how to construct a pattern matrix, we'll study the 2-input AND function (called  $x \wedge y$  in Table 1), which maps  $(x, y) \in \mathbb{F}_2^2$  to  $xy \in \mathbb{F}_2$ . (This function has  $n = 2$  inputs and  $m = 1$  output, so its pattern matrix dimensions are  $2^{2^m} = 2^{2^1} = 4$  by  $2^{2^n} = 2^{2^2} = 16$ .)

We'll start with the simplest possible pattern basis,  $\mathcal{P}_{\text{ind}}$ ; this uses the indicator basis, which is also the canonical basis, in each property space. This will give us a trivial pattern matrix (shown below in (1)), and a measured complexity of 0 for every function, but it makes the pattern matrix easy to understand (and its basis can be changed later).

Using this basis, each row of the pattern matrix  $\text{PM}_{\mathcal{P}_{\text{ind}}}(f)$  corresponds to an  $m$ -input 1-output truth table (or an “ $(m, 1)$ -table” for short). We'll order the rows lexicographically, so for  $m = 1$  they are labelled by: the constant function 0 (with one input, which it ignores), the identity function (sometimes called a “wire”), the NOT function, and the constant function 1. (For convenience, we'll often identify a truth table with the function it describes.)

Each column of this pattern matrix corresponds to one  $n$ -input 1-output truth table (or “ $(n, 1)$ -table”), of which there are  $2^{2^n} = 16$ . We'll put these in “doubly-lexicographic order” — the possible values of the 2 inputs  $(x, y)$  will be ordered  $(0, 0), (0, 1), (1, 0), (1, 1)$ , which tells us how to arrange each truth table's 4 output bits as a 4-tuple; the 16 possible 4-tuples will be sorted lexicographically to label the columns, giving a column order of  $(0, 0, 0, 0), (0, 0, 0, 1), (0, 0, 1, 0), (0, 0, 1, 1), \dots (1, 1, 1, 0), (1, 1, 1, 1)$ . (The AND function itself has a truth table of  $(0, 0, 0, 1)$  in this format, so it labels the 2nd column in this order.)

To construct the indicator-basis pattern matrix for AND (or for any function  $f$ ), we compose its output with each  $(m, 1)$ -table  $t_o$  in turn, and see which  $(n, 1)$ -table  $t_i$  is induced on its inputs — i.e., which  $(n, 1)$ -table describes the composition  $t_o \circ f$ . This tells us which matrix entry is 1 (with all others being 0) in the row labelled by  $t_o$ . (This is a kind of “pullback” operation on  $f$ , as we'll discuss when the pattern matrix is defined formally in Section 4.3.)

In the special case of  $m = 1$ , the 2nd row is labelled by  $t_o = (0, 1)$  (which describes the identity function on  $\mathbb{F}_2$ ), so that row just indicates  $f$ 's ordinary truth table (by putting its single 1 in the corresponding column). For the AND function, whose truth table is  $(0, 0, 0, 1)$ , that row gets a 1 in the 2nd column (as can be seen in the full matrix shown below).

The 3rd row is labelled by  $t_o = (1, 0)$  (the NOT function), so it just indicates the truth table of  $\text{NOT}(f)$  (in the same manner). The 1st and 4th rows (labelled by a



$t_o$  of  $(0,0)$  and  $(1,1)$  respectively, which describe constant functions) are the same for all functions (provided we're using this basis  $\mathcal{P}_{\text{ind}}$  and  $m = 1$ ); they indicate the all-0 and all-1  $n$ -input truth tables respectively, by placing their single 1 entry in the first and last columns respectively.

For the AND function, we end up with the following indicator-basis pattern matrix (which differs from AND's pattern matrix in Table 1, because we used a different pattern basis there):

$$A = \text{PM}_{\mathcal{P}_{\text{ind}}}(\text{AND}) = \begin{vmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{vmatrix} \quad (1)$$

Of course this pattern matrix is not a very *compact* encoding of the AND function! But that's not its purpose — what interests us is (1) its matrix multiplication corresponds to function composition; (2) after suitable changes of basis, its entries can tell us the relationship between *any* properties of  $t_o$  and  $t_i$ , in any circuit which computes  $t_i$  using the structure  $t_o \circ f$  — that is, any circuit whose first stage is  $f$  and whose remainder is described by  $t_o$ . This lets the pattern matrix *linearly* encode everything about  $f$ 's possible role in any 1-output circuit's computation (provided  $f$  has been “widened” to include whatever wires run alongside it in the circuit, as described earlier). And feature (1) makes it possible to use this matrix (in *any* basis) to define subadditive measures with respect to function composition (which any generally useful circuit complexity lower bound ought to be).

(The fourier matrices mentioned earlier also have feature (1) from that list — indeed, they *can* be used to define subadditive measures on boolean functions — but they lack feature (2), which is part of the reason the subadditive measures they help define can never have nontrivially large values. A more basic reason is just their smaller size, though this could be seen as a different aspect of the same limitation.)

Of course, none of this will be interesting until we *do* change that matrix's basis (which requires *finding* a useful basis to change it to, as we've discussed elsewhere). For now, we'll just show how we got the pattern matrices in Table 1, by changing from the indicator basis to  $\mathcal{P}_{\text{Rm2b}}$ , which is “useful” for  $n = 2$ . (Whether  $\mathcal{P}_{\text{Rm2b}}$  is usefully extendable to higher  $n$  is not yet known.)

### 2.17.4 change of pattern basis

First we should describe our conventions for change of basis matrices (which follow [MB99]). A change of basis matrix  $P$  should consist of columns giving each “old” basis element in “new” coordinates (i.e. in coordinates relative to the new basis). A column vector  $v$ , using old coordinates, can then be changed to a vector  $v'$ , using new coordinates, by

$$v' = A v$$

and a square matrix  $A$ , represented in the old basis, can be changed to  $A'$ , expressing the same linear map relative to the new basis, using

$$A' = P A P^{-1}$$

Since a pattern matrix is in general rectangular, we need to use families of bases (one for each possible matrix dimension). By an abuse of notation, within a pattern basis  $\mathcal{P}$  we'll let  $P_n = \mathcal{P}_{P_n}$  (for each  $n \geq 0$ ) denote not only a basis set  $P_n \subset Q_n$  (of the property space  $Q_n = \mathcal{P}_{Q_n}$ ), but also the change of basis matrix from the indicator basis to that basis set (ordering its rows and columns in a standard way implied by the context).<sup>30</sup> If  $A$  is a pattern matrix representing the pattern map of  $f: \mathbb{F}_2^n \rightarrow \mathbb{F}_2^m$  using the indicator pattern basis (as in (1)), we can change it to a different pattern matrix  $A'$ , representing the same pattern map using a different pattern basis  $\mathcal{P}$  (containing the property space bases (and change of basis matrices)  $\langle P_n \rangle$ ), by

$$A' = P_m A P_n^{-1}$$

In the notation we used earlier, in which the pattern basis used by a pattern matrix is indicated by a subscript on PM (so the pattern matrices  $\text{PM}_{\mathcal{P}}(f)$  and  $\text{PM}_{\mathcal{P}_{\text{ind}}}(f)$  use the pattern bases  $\mathcal{P}$  and  $\mathcal{P}_{\text{ind}}$  respectively), we can express that instead as

$$\text{PM}_{\mathcal{P}}(f) = P_m \text{PM}_{\mathcal{P}_{\text{ind}}}(f) P_n^{-1} \quad (2)$$

---

<sup>30</sup> This overloaded meaning for the notation  $P_n$  is compatible with “tensoring basis sets” (i.e., forming a basis for a tensor product of vector spaces  $V_1 \otimes V_2$ , by tensoring every pair of elements of existing bases for  $V_1$  and  $V_2$ ) — for example,  $P_n \otimes P_n$  has a compatible meaning, whether we interpret  $P_n$  as a basis set, or as that basis set’s change of basis matrix from the canonical (indicator) basis. This also explains the notation  $P_{n+1} = W_{n+1}(P_n \otimes P_n)$  for the “appropriate change of basis” (in an inductive construction of the basis family  $\langle P_n \rangle$ ) which we discussed in Section 2.13.

Thus, to construct  $\text{PM}_{\mathcal{P}_{\text{Rm2b}}}(\text{AND})$  as shown in Table 1, we have:

[illegible]

(Note that  $P_2^{-1}(\text{Rm2b}) = P_2^{(\text{Rm2b})}$ , since it's orthonormal and symmetric.<sup>31</sup>)

### 2.17.5 comparing choices of pattern basis $\mathcal{P}$ and matrix measure $M$

Finally, we compare the measured complexity given by several choices of pattern basis  $\mathcal{P}$  and matrix measure  $M$ , noting that it's always subadditive (with respect to function composition), even though it's not always useful, or related to actual circuit complexity.

In the following table, we compare several variants of  $C_{M,\mathcal{P}}$  (and two kinds of actual complexity), for a few simple functions with 1 or 2 outputs and 0 to 2 inputs, including the functions that were shown in Table 1. In the first three entries of each numeric column, the reader can verify subadditivity of that column’s complexity measure, since the first row’s function  $(\bar{x} \wedge y)$  is the composition of the functions in the next two rows  $((x \wedge y)$  and  $(\bar{x}, y))$ . Other observations about the tabulated values are discussed below.

<sup>31</sup> In general, a pattern basis change of basis matrix is not necessarily even orthogonal (or unitary), let alone also symmetric; even when it can be expressed that way, this depends on the arbitrary order and sign (or complex phase) of the basis elements. But in all our listed examples (except  $\mathcal{P}_{\text{rand1}}$ ), it's either already true, or can be made true by rescaling some patterns by roots of unity.

**Table 2:** value of  $C_{M,\mathcal{P}}(f)$ , for various functions  $f$  and choices of  $M$  and  $\mathcal{P}$ , compared to two kinds of actual complexity

	choice of $\mathcal{P}$ :	$\mathcal{P}_{\text{mon}}$		$\mathcal{P}_{\text{Rm2b}}$		$\mathcal{P}_{\text{rand1}}$		or, kind of actual complexity:	
	choice of $M$ :	$M_{\text{nz}}$	$M_{\text{abs}}$	$M_{\text{nz}}$	$M_{\text{abs}}$	$M_{\text{nz}}$	$M_{\text{abs}}$	$C_{B_2}$	“free XORs”
2-input functions:	(function nickname)								
$(\bar{x} \wedge y)$	AND( $\bar{x}, y$ )	2	1	2	1	4	1.807	1	1
$(x \wedge y)$	AND	2	1	2	1	4	1.840	1	1
$(\bar{x}, y)$	<i>complement first input</i>	0	0	0	0	4	1.812	0	0
$(x \oplus y, y)$	<i>controlled-NOT</i>	0	0	0	0	4	1.772	1	0
$(x \oplus y)$	XOR	2	1	0	0	4	1.834	1	0
$(x)$	<i>discard 2nd input</i>	2	1	0	0	4	1.846	0	0
1-input $f$ s:									
$(x, x)$	<i>splitter</i>	0	1	0	0	2	1.385	0	0
$(\bar{x})$	NOT	0	0	0	0	2	0.917	0	0
0-input $f$ s:									
$(0)$	<i>create 0</i>	0	$\frac{1}{2}$	0	0	1	0.966	0	0

The functions are represented as tuples of output formulae (as in Table 1), in which the boolean operations AND, XOR, and NOT are represented by  $\wedge$ ,  $\oplus$ , and  $\bar{x}$  respectively. The 0-input function shown as  $(0)$  has one output, which is always  $0 \in \mathbb{F}_2$ . The pattern bases  $\mathcal{P}_{\dots}$  are defined in Section 2.17.2; we don’t show  $\mathcal{P}_{\text{ind}}$  here, since its measured complexity is 0 for every function (for either choice of  $M$ ).

As in Table 1, the matrix measures  $M_{\text{nz}}$  and  $M_{\text{abs}}$  measure (respectively) the maximum number of nonzero entries, or the maximum sum of absolute values of entries, in any pattern matrix row. (Their necessary properties are proven in Section 4.)

The kinds of actual complexity shown are  $C_{B_2}$  (all 0- or 1-input gates are free;

all 2-input 1-output gates have cost 1) [Wil11]<sup>32</sup>, and “free XORs” or “multiplicative complexity” (the same, except  $\mathbb{F}_2$ -affine gates (like XOR) are also free) [BPP00].

About the measured complexity ( $C_{M,\mathcal{P}}$ ) values themselves, we note:

- The random pattern basis  $\mathcal{P}_{\text{rand1}}$  sees similarly high “complexity” in every function (given its number of inputs), as predicted in Sections 2.8 and 2.9; the only exceptions are identity functions (not shown in the table), such as those computed by some circuits consisting only of wires.
- The monomial basis  $\mathcal{P}_{\text{mon}}$  (when used with  $M_{\text{nz}}$ ) is measuring “amount of lack of invertibility” (which is also subadditive), rather than actual complexity. (The fact that it does this was mentioned in Section 2.9, and is easy to prove. When used with  $M_{\text{abs}}$  instead,  $\mathcal{P}_{\text{mon}}$  also sees positive “measured complexity” in invertible functions which are not onto.)
- The “candidate useful” basis  $\mathcal{P}_{\text{Rm2b}}$  agrees with the “free XORs” kind of actual complexity (up to a constant factor). (This is not evidence that the “free XORs” kind of complexity is more natural — rather, it reflects the author’s preexisting guess that it is, since measuring  $\mathbb{F}_2$ -affine functions as having 0 complexity was built into the definition of “niceness” which guided the search for that basis.)
- There is not yet any evidence about whether  $M_{\text{nz}}$  or  $M_{\text{abs}}$  is “better”.<sup>33</sup>

---

<sup>32</sup> Our definition of  $C_{B_2}$  differs from that reference in a trivial way, which is to give isolated NOT gates 0 cost. This makes no difference except for a function output which is the logical negation of an input variable, since all other NOT gates can be absorbed into connected gates without increasing their cost. We do this to simplify the behavior of cost when circuits are composed.

<sup>33</sup> On the other hand, we can rule out using an  $M$  defined as the maximum *euclidean* norm of any pattern matrix row, since in any orthonormal pattern basis it would see all reversible functions as having 0 complexity.

### 3 Notation

$\text{aw}_k(g)$	the “adjoin wires operation” (for adjoining $k \geq 0$ non-interacting wires in parallel) applied to the circuit $g$ ; can also be applied to a function $f$ ; commutes with $f_{\text{circ}}$ (defined in Section 4.5)
$b$	any “bit” $b \in \mathbb{F}_2$ ; sometimes represented by $(-1)^b \in \mathbb{R}$
$\mathbb{C}$	the field of complex numbers
$C$	any “cost” or “complexity” measure (with values in $\mathbb{R}^{\geq 0}$ ) which obeys an appropriate kind of “subadditivity”; also used occasionally for a matrix product $C = AB$ , or an arbitrary circuit $C$
$C_{\text{cost}}(g)$	circuit cost (of a circuit or gate $g$ ), as defined by a circuit model determined by the context
$C_{\text{min}}(f)$	circuit complexity (of a function $f$ ), defined as the minimum cost of any circuit (in a given model) which computes $f$
$C_K$	Kolmogorov complexity
$C_{M,\mathcal{P}}$	a specific (unscaled) <b>measured complexity formula</b> , defined (below) using a submultiplicative matrix measure $M$ and a pattern basis $\mathcal{P}$
$C_{M,\mathcal{P},\lambda}$	a scaled version of $C_{M,\mathcal{P}}$ , defined as $\lambda C_{M,\mathcal{P}}$ for some constant $\lambda > 0$
$\text{DualPatMap}_{\mathcal{P}}$	the <b>dual pattern map</b> (defined below) of its argument $f$ with respect to pattern basis $\mathcal{P}$ ; used in the definition of $\text{PM}_{\mathcal{P}}(f)$
$f$	a function (typically boolean, $f: \mathbb{F}_2^n \rightarrow \mathbb{F}_2^m$ ) whose circuit complexity we’re interested in
$f_{\text{circ}}(g)$	the function computed by the circuit or gate $g$
$\mathbb{F}_q$	the finite field with $q$ elements (mainly used for its special case, $\mathbb{F}_2$ )
$\mathbb{F}_q^n$	an $n$ -dimensional (finite) vector space over $\mathbb{F}_q$
$\mathcal{G}$	the set of <b>generating circuits</b> , in a circuit model under discussion

(continued on next page)

(*notation, continued*)

$\mathcal{G}_{\text{core}}$	the finite set of “small” generating circuits on which most of the others are based, when assuming $\mathcal{G}$ has a certain “conventional” structure (defined in Section 4.5)
$GL(n, K)$	the “general linear group” consisting of automorphisms of the vector space $K^n$
$K$	any field (though some of our definitions and proofs could still work if it was a ring)
$K[S]$	the free vector space on the finite set $S$ over the field $K$ ; its elements are the formal linear combinations of elements of $S$ (with coefficients in $K$ )
$K^S$	the algebra (or vector space) of arbitrary $K$ -valued <b>properties</b> (functions) over the finite set $S$ , with elementwise addition and multiplication; identified with $K[S]^*$ (the vector-space dual of $K[S]$ )
$L_{1,\infty}$	the submultiplicative matrix norm defined by $\ A\ _{1,\infty} = \max_j \sum_i  a_{ij} $
$m$	(often used as the number of outputs (0 or more) of a function $f$ )
$m_g$	the number of outputs of a circuit or gate $g$
$M$	any submultiplicative matrix measure
$M_{\text{nz}}(A)$	for any matrix $A$ , the maximum number of nonzero entries on any row ( $M_{\text{nz}}$ is a submultiplicative matrix measure, as we prove below)
$M_{\text{abs}}(A)$	for any matrix $A$ , the maximum sum of absolute values of entries on any row (also a submultiplicative matrix measure)
$\text{Maps}(S, T)$	the set of arbitrary functions from the elements of $S$ to the elements of $T$ (ignoring any structure $S$ or $T$ might have); if $S$ or $T$ is a vector space, all its vectors are “elements” (notation borrowed from [Zag04])
$\text{Mats}_{\geq 1}(K)$	the “partial ring” of all matrices with entries in $K$ and each dimension at least 1 (note that whenever matrices’ dimensions allow them to be added or multiplied, they obey all ring axioms)
$n$	(often used as the number of inputs (0 or more) of a function $f$ )

(*continued on next page*)

(*notation, continued*)

$n_g$	the number of inputs of a circuit or gate $g$
NP	the complexity class of languages nondeterministically recognized by polynomial-time algorithms
$O, \Omega, \Theta$	notations for asymptotic limits (using the definitions from computer science, not analytic number theory); see e.g. [AB09] and [WikO]
$\tilde{O}$	$\tilde{O}(X)$ is an upper bound like $O(X)$ , but ignores logarithmic factors in $X$
P	the complexity class of languages recognized by polynomial-time algorithms
P/poly	the complexity class of languages recognized by (non-uniform) boolean circuits of polynomial size
$\mathcal{P}$	any <b>pattern basis</b> (defined below); see Section 2.17.2 for definitions of a few specific pattern bases, such as $\mathcal{P}_{\text{ind}}$ , $\mathcal{P}_{\text{mon}}$ , or $\mathcal{P}_{\text{Rm2b}}$
$P_n$	the special basis of “patterns” in the property space $Q_n$ (in a pattern basis $\mathcal{P}$ ); the same notation can indicate a basis set or its change of basis matrix
$P_n^{(\text{Rm2b})}$	a parenthesized superscript on $P_n$ indicates it’s part of the named pattern basis (in this case, $\mathcal{P}_{\text{Rm2b}}$ )
$\text{PatMap}_{\mathcal{P}}$	the <b>pattern map</b> (defined below) of its argument $f$ with respect to pattern basis $\mathcal{P}$ ; used in the definition of $\text{PM}_{\mathcal{P}}(f)$
$\text{PM}_{\mathcal{P}}(f)$	the <b>pattern matrix</b> (defined below) of function $f$ , with respect to pattern basis $\mathcal{P}$
$\mathcal{Q}$	the operation to construct each $Q_n$ (in a pattern basis $\mathcal{P}$ )
$Q_n$	the <b>property space</b> for input or output tuples of size $n$ (in a pattern basis $\mathcal{P}$ )
$\mathbb{R}$	the field of real numbers
$\mathbb{R}^{\geq 0}$	the nonnegative reals (notation borrowed from [Odo14])
$\mathcal{T}$	the operation to construct each $T_n$ (in a pattern basis $\mathcal{P}$ )

(*continued on next page*)



(*notation, continued*)

- $T_n$  the **configuration space** for input or output tuples of size  $n$  (in a pattern basis  $\mathcal{P}$ )
- $V_n$  the **value space** for input or output tuples of size  $n$  (in a pattern basis  $\mathcal{P}$ )
- $X^*$  the dual of  $X$
- $\oplus$  direct sum, or logical XOR (according to context); but in some contexts XOR is represented instead as  $+$  (e.g. for vectors over  $\mathbb{F}_2^n$ ) or as multiplication (e.g. for bits  $b \in \mathbb{F}_2$  represented by  $(-1)^b \in \mathbb{R}$ )
- $\otimes$  tensor product
- $\wedge$  logical AND, in boolean formulas
- $\bar{x}$  logical negation of  $x$ , in boolean formulas
- $\bar{x}$  complex conjugate of  $x$  (looks similar to logical negation, but the context can disambiguate them)

## 4 Definitions and Theorems

For convenience, we repeat a few of the definitions from earlier sections, since they're used directly in the proofs below. For others, refer to Section 2.

In the following,  $A$  and  $B$  refer to matrices with entries (in any ring  $K$ )  $a_{ij}$  and  $b_{jk}$  respectively. When their matrix product is needed, it's a matrix  $C = AB$  with entries  $c_{ik} = \sum_j a_{ij}b_{jk}$ . (We also use  $C$  to denote an unspecified “subadditive” matrix measure, but the context should make each usage clear.)

### 4.1 matrix measures

By a **submultiplicative matrix measure**, we mean any  $M: \text{Mats}_{\geq 1}(K) \rightarrow \mathbb{R}^{\geq 0}$  for which (for all matrix products  $AB$ )

$$M(AB) \leq M(A)M(B)$$

By a **“subadditive” matrix measure**, we mean any  $C: \text{Mats}_{\geq 1}(K) \rightarrow \mathbb{R}^{\geq 0}$  for which (for all matrix products  $AB$ )

$$C(AB) \leq C(A) + C(B)$$

Note that the left side is *not*  $C(A + B)$ ; this is an “abuse of notation” justified by thinking of function composition (and thus the matrix multiplication which represents it) as analogous to addition, as explained where this terminology was earlier defined (Section 2.5).

We will also speak of a “***subadditive***” ***measure*** more generally — for example, by a “***subadditive***” ***measure with respect to function composition*** we mean any  $C$  for which

$$C(f \circ g) \leq C(f) + C(g)$$

**Definition (nz):**  $\text{nz}: K \rightarrow \mathbb{R}^{\geq 0}$  is for counting nonzero entries:

$$\text{nz}(x) = \begin{cases} 0, & \text{if } x = 0; \\ 1, & \text{if } x \neq 0. \end{cases}$$

**Remark:**  $\text{nz}$  (in any ring  $K$ ) is subadditive and submultiplicative. (Note that the comparisons in those definitions are done in  $\text{nz}$ ’s codomain  $\mathbb{R}^{\geq 0}$ , so there is no need for  $K$  to be ordered.)

**Definition ( $M_{\text{nz}}$ ):** For any matrix  $A$ ,  $M_{\text{nz}}(A)$  is defined as the maximum number of nonzero entries on any row:<sup>34</sup>

$$M_{\text{nz}}(A) = \max_i \sum_j \text{nz}(a_{ij})$$

**Remark:**  $M_{\text{nz}}(A) \geq 1$ , unless  $A$  is a zero matrix.

**Theorem 1**  $M_{\text{nz}}(A)$  is a submultiplicative matrix measure.

---

<sup>34</sup>  $M_{\text{nz}}$  is related to the so-called “zero norm” (which counts nonzero coordinates of a vector), though neither of them is actually a norm.

**Proof.** For  $C = AB$ , we have

$$\begin{aligned}
M_{\text{nz}}(C) &= \max_i \sum_k \text{nz}(c_{ik}) \\
&= \max_i \sum_k \text{nz} \left( \sum_j a_{ij} b_{jk} \right) \\
&\leq \max_i \sum_k \sum_j \text{nz}(a_{ij} b_{jk}) && \text{(since nz is subadditive)} \\
&\leq \max_i \sum_k \sum_j (\text{nz}(a_{ij}) \text{nz}(b_{jk})) && \text{(since nz is submultiplicative)} \\
&= \max_i \sum_j \left( \text{nz}(a_{ij}) \sum_k \text{nz}(b_{jk}) \right) \\
&\leq \max_i \sum_j \left( \text{nz}(a_{ij}) \max_j \sum_k \text{nz}(b_{jk}) \right) && \text{(ok since nz}(a_{ij}) \text{ is nonnegative)} \\
&= \max_i \sum_j \left( \text{nz}(a_{ij}) M_{\text{nz}}(B) \right) \\
&= M_{\text{nz}}(B) \max_i \sum_j \left( \text{nz}(a_{ij}) \right) \\
&= M_{\text{nz}}(B) M_{\text{nz}}(A)
\end{aligned}$$

Combining,

$$M_{\text{nz}}(AB) \leq M_{\text{nz}}(A) M_{\text{nz}}(B)$$

■

**Remark, and definition of  $M_{\text{abs}}$ :** The preceding proof used only the properties of  $\text{nz}$  named in the proof's annotations; thus for any other subadditive and submultiplicative  $N: K \rightarrow \mathbb{R}^{\geq 0}$  used in place of  $\text{nz}$ , we could define a similar matrix measure  $M_N$  and prove it submultiplicative. For example, if  $K = \mathbb{R}$  and  $N(x) = |x|$ , we get the other matrix measure used in the tables in Section 2.17, namely  $M_{\text{abs}}(A) = M_N(A) = \|A^T\|_{1,\infty}$ , where  $\|\cdot\|_{1,\infty}$  denotes the (submultiplicative)  $L_{1,\infty}$  matrix norm.<sup>35</sup>

---

<sup>35</sup> To use  $M_{\text{abs}}$  in  $C_{M,\mathcal{P}}$  below, we'll also need to prove it's at least 1 for any pattern matrix we apply it to. We only use it with pattern bases which are orthonormal in the standard inner product on the property space; the change of basis from the indicator-basis pattern matrix to the actual one thus preserves the sum of squares of absolute values of matrix entries, so at least one pattern matrix row has that sum at least 1. If any entry has absolute value more than 1, we're

## 4.2 pattern basis (and possible generalizations)

We’ll define a pattern basis somewhat more generally here than in the rest of the paper, since the more general version might be useful or illuminating, and works just as easily in our proofs. Below, we’ll spell out how most of our discussions have assumed a specific kind of pattern basis, but we’ll point out some generalizations (most of which fit within this definition — the others could be easily added) which might be fruitful to explore.

**Definition (pattern basis):**

A *pattern basis*  $\mathcal{P} = (\langle V_n \rangle, \mathcal{T}, \mathcal{Q}, \langle P_n \rangle)$  embodies:

- a family of “value spaces”  $\langle V_n \rangle$ , one for each possible input or output space of any function  $f$  whose circuit complexity is to be lowerbounded (or which will be composed to make such functions).
- an operation  $\mathcal{T}$ , which turns each value space  $V_n$  into an associated “configuration space”  $T_n$ , which must be a space of maps *from*  $V_n$ ; the only restriction on what its elements map *to* (and on exactly which maps it contains) is that, for this pattern basis to be **applicable** to a given circuit model, the model’s constructible functions  $f: V_n \rightarrow V_m$  must be composable with any element  $t_o$  of  $T_m$  to get an element  $t_o \circ f$  of  $T_n$ .
- an operation  $\mathcal{Q}$ , which turns each  $T_n$  into the associated “property space”  $Q_n = K[T_n]^* = K^{T_n}$ , an algebra (and vector space) over a fixed field  $K$ , whose elements are arbitrary  $K$ -valued “properties” (functions) of  $T_n$  (as a set), with elementwise addition and multiplication;
- a choice of vector-space basis  $P_n$  for each  $Q_n$ .

(Given a pattern basis  $\mathcal{P}$ , we can refer to these objects by  $\mathcal{P}_{\langle V_n \rangle}$  (for the family) or  $\mathcal{P}_{V_n}$  (for individual  $n$ ),  $\mathcal{P}_{\mathcal{T}}$ ,  $\mathcal{P}_{\mathcal{Q}}$ , and  $\mathcal{P}_{\langle P_n \rangle}$  or  $\mathcal{P}_{P_n}$ .)

In most of the present paper (outside this section), when discussing any pattern basis we assume for simplicity that:

- its  $V_n$  is always  $\mathbb{F}_2^n$  for some  $n \geq 0$  (i.e. it’s about *boolean* functions rather than some other kind);

---

done; otherwise each row’s sum of absolute values is at least its sum of their squares, completing the proof.

- its  $\mathcal{T}$  turns  $V_n$  into the set of “ $n$ -bit truth tables”  $T_n = \text{Maps}(V_n, \mathbb{F}_2)$  (which clearly meets the condition about compositions existing, for any boolean circuit model).

(Thus its property spaces  $Q_n$  will be vector spaces over  $K$  of dimension  $2^{2^n}$ .)

But other choices would also be interesting to explore:

- We might use  $\mathbb{F}_q^n$  (rather than  $\mathbb{F}_2^n$ ) for  $V_n$ , thus studying general functions with inputs and outputs having value-sets  $\mathbb{F}_q$ , rather than only boolean functions.
- (We could even consider  $V_n$  which are direct products of distinct value-sets, potentially different for each input or output, and not necessarily finite fields (since they need no structure to work in these definitions).)
- We might use  $\mathbb{F}_q$  (or a structureless set of  $q$  elements) in place of  $\mathbb{F}_2$ , as the codomain of the maps in the space  $T_n$  — that is, use “truth tables” whose entries each have  $q$  possible symbols rather than 2. (In principle this is an independent choice from the value-sets used to make  $V_n$ , though there may be advantages in keeping them the same. Other than that, I know of no advantage to using more than 2 symbols.)
- If we’re especially interested in *reversible* (i.e. bijective) functions (and choose our circuit model accordingly), we can assume  $n = m$  in  $f: V_n \rightarrow V_m$ , which makes the composition condition less restrictive. In particular, this would let us define  $T_n$  as the set of functions taking  $V_n$  to itself (i.e. the set of permutations on its elements — a richer object than the truth tables taking it to  $\mathbb{F}_2$ ), which is the same as the set of functions we’re interested in. That set of functions (for each  $n$ ) is in fact a group under composition (which we can call  $G_n$ ), so  $Q_n^* = K[T_n]$  is the *group algebra*  $K[G_n]$  (but note that its product is based on the group operation of function composition, *not* on the elementwise product normally used in its vector-space dual  $Q_n = K^{T_n}$ ). The “dual pattern map” of a function  $f$  (which we’re about to define below) then turns out to be an element of the *regular representation* of  $G_n$  acting on  $Q_n^*$ . This suggests looking to the rich theory of *finite group representations* (see e.g. [WikR], [Seg14], [VO05], [Web16], [Zag04]) as a possible source of insights for constructing a useful pattern basis. (It may provide insights even for arbitrary functions and the usual choice of  $Q_n$ , since the behavior of the reversible functions (especially, those which should have 0 measured complexity) still puts important constraints on the pattern basis.)

- We could instead restrict attention to some subset  $\mathcal{F}_n$  (closed under composition and inverse) of the  $n$ -input reversible functions. This will be a subgroup of the  $G_n$  considered above, and remains amenable to the same kind of group-representation-based construction. One example is  $GL(n, \mathbb{F}_q)$ , the group of invertible  $\mathbb{F}_q$ -linear transformations of  $\mathbb{F}_q^n$ . These have a maximum circuit complexity of approximately  $\log_q q^{n^2} = n^2$ , but as with general boolean functions (at least when  $q = 2$ ), circuits of complexity  $\tilde{O}(n)$  (and possibly  $O(n)$ ) can compute functions which “look random” within this class [Aar07], and no explicit function families have been proven to have superlinear complexity in terms of  $n$  [Aar07a], though specific functions can be defined which intuitively seem almost certain to be more complex [Coh00].
- Still restricting attention to reversible functions, we could instead make our usual choice of  $T_n$  *smaller* — such functions preserve density of its truth tables, so we could restrict it to truth tables of a single fixed density; at least intuitively, for non-extreme densities this seems likely to retain most information about complexity. (It certainly retains the distinction between individual functions.) Similarly (if we generalized our definition further to permit other constructions of  $Q_n$ ), we could make  $Q_n$  smaller, by restricting it to properties which (as polynomials in truth table entries represented by  $(-1)^b$ ) are homogeneous of a given degree, since the degree of their terms is also preserved by reversible functions.

When comparing different possible  $\mathcal{P}$  for one circuit model, it often happens that most objects in  $\mathcal{P}$  are fixed by the context, but not the choice of basis  $P_n$  for each  $Q_n$ ; in that case (as a notational convention) we might call  $\langle P_n \rangle$  alone “the pattern basis”, or call one possible choice for it “a pattern basis”, and use it in place of  $\mathcal{P}$  in notations which logically depend on the entire pattern basis, like the measured complexity formula  $C_{M, \mathcal{P}}$  (defined below). We might also refer to  $P_n$  (for a single value of  $n$ ) as a “pattern basis for  $Q_n$ ”. These conventions are meant to make it simpler to discuss or inductively construct various choices for  $\langle P_n \rangle$  which might make  $\mathcal{P}$  “useful” for a given circuit model.

Note also that many of the objects in  $\mathcal{P}$  might be treated as having more or less structure for different purposes. For example,  $V_n$  and  $T_n$  are treated as sets by the present definitions, but can be treated as having more structure when convenient for defining operations on them (e.g. when adjoining a wire to a circuit). Similarly,  $Q_n$  needs only vector space structure to let us *define* the term “pattern basis”, but a typical scheme for *constructing* a pattern basis is likely to use  $Q_n$  as an algebra (and

to use whatever structure is available in the associated  $Q_n^*$ ,  $T_n$ , and  $V_n$  as well). Some schemes for constructing the basis sets  $P_n$  may find it useful to give them additional structure too (such as a group structure).

### 4.3 pattern matrix

#### Definition (pattern matrix):

With respect to a pattern basis  $\mathcal{P}$ , we can construct the **pattern matrix**  $\text{PM}_{\mathcal{P}}(f)$  for any function  $f: V_n \rightarrow V_m$ , provided  $V_n$  and  $V_m$  are in  $\mathcal{P}_{\langle V_n \rangle}$ , and  $f$  is constructible in a circuit model to which  $\mathcal{P}$  is applicable. (This construction is described informally in Section 2.6, and illustrated in Sections 2.17.3 and 2.17.4.)

The pattern matrix describes how  $f$  transforms patterns between the property spaces  $Q_n$  and  $Q_m$  associated with  $f$ 's input and output spaces (i.e. with  $V_n$  and  $V_m$ ); its matrix entries can be directly used in the measured complexity formula  $C_{M,\mathcal{P}}$  which we'll define shortly.

(In the following definition we assume  $\mathcal{P} = (\langle V_n \rangle, \mathcal{T}, \mathcal{Q}, \langle P_n \rangle)$ , and that  $n$  and  $m$  are chosen so that  $f: V_n \rightarrow V_m$ .)

First we construct the associated configuration spaces

$$T_n = \mathcal{T}(V_n)$$

$$T_m = \mathcal{T}(V_m)$$

and property spaces

$$Q_n = \mathcal{Q}(T_n)$$

$$Q_m = \mathcal{Q}(T_m)$$

Recall from the definition of “pattern basis” that for any  $t_o \in T_m$  and any constructible  $f$  (in a circuit model  $\mathcal{P}$  is applicable to), we require  $t_o$  to be a function on  $V_m$  which can be composed with  $f$  to make  $t_i = t_o \circ f$ , with  $t_i \in T_n$ . (In other words,  $f$  can “pull back”  $t_o: V_m \rightarrow X$  to get  $t_i: V_n \rightarrow X$ , for some  $X$ . The notations  $t_o$  and  $t_i$  come from  $t_o$  being a truth table applied to  $f$ 's outputs, and  $t_i$  being the induced truth table applicable to  $f$ 's inputs.)

The resulting map from  $T_m$  to  $T_n$  (i.e. the “pullback by  $f$ ”) can then be “**linearized**” to map the free vector space over  $K$  with the basis  $T_m$  (treated as a set, ignoring any vector space structure it might have), i.e.  $K[T_m] = Q_m^*$ , into the similarly defined  $Q_n^*$ ; we call the result the **dual pattern map** of  $f$  (with respect to  $\mathcal{P}$ ). (In Section 6.3, we'll discuss what we gain by doing this linearization.)

To make the definition concise, we'll use  $t_o \in T_m$  as both an index of the canonical basis of  $K[T_m]$  and as the corresponding basis element; an arbitrary dual property  $t \in Q_m^*$  can then be uniquely expressed as  $\sum_{t_o \in T_m} a_{t_o} t_o$  with coefficients  $a_{t_o} \in K$ , and the dual pattern map takes it to a corresponding sum in  $Q_n^*$ :

$$\text{DualPatMap}_{\mathcal{P}}(f): Q_m^* \rightarrow Q_n^*$$

$$\text{DualPatMap}_{\mathcal{P}}(f) \left( \sum_{t_o \in T_m} a_{t_o} t_o \right) = \sum_{t_o \in T_m} a_{t_o} (t_o \circ f)$$

(noting that  $t_o \circ f \in T_n$  as discussed above, and treating it as the corresponding canonical basis element of  $K[T_n] = Q_n^*$ ). Note that  $\text{DualPatMap}_{\mathcal{P}}(f)$  is clearly linear, with respect to the vector space structure (over  $K$ ) of  $Q_m^*$  and  $Q_n^*$ .

Finally (and in a sense optionally — working directly with dual patterns might turn out to be just as convenient), we dualize  $\text{DualPatMap}_{\mathcal{P}}(f)$  to get the **pattern map** of  $f$ :

$$\text{PatMap}_{\mathcal{P}}(f): Q_n \rightarrow Q_m$$

$$\text{PatMap}_{\mathcal{P}}(f) = (\text{DualPatMap}_{\mathcal{P}}(f))^*$$

and then define the **pattern matrix**  $\text{PM}_{\mathcal{P}}(f)$  as simply the matrix which represents this linear map, with respect to the bases  $P_n$  of  $Q_n$  and  $P_m$  of  $Q_m$  (both taken from  $\mathcal{P}_{\langle P_n \rangle}$ ). ■

**Theorem 2** *Pattern matrices multiply, and pattern maps compose, like their functions compose. That is, if  $\mathcal{P}$  is applicable to functions  $f$  and  $g$  (and if  $f$ 's domain matches  $g$ 's codomain, so  $f \circ g$  exists), then  $\mathcal{P}$  is applicable to  $f \circ g$ , and:*

$$\text{DualPatMap}_{\mathcal{P}}(f \circ g) = \text{DualPatMap}_{\mathcal{P}}(g) \circ \text{DualPatMap}_{\mathcal{P}}(f) \quad (3)$$

$$\text{PatMap}_{\mathcal{P}}(f \circ g) = \text{PatMap}_{\mathcal{P}}(f) \circ \text{PatMap}_{\mathcal{P}}(g) \quad (4)$$

$$\text{PM}_{\mathcal{P}}(f \circ g) = \text{PM}_{\mathcal{P}}(f) \text{PM}_{\mathcal{P}}(g) \quad (5)$$

**Proof.** Let  $\mathcal{P}$  contain  $\langle V_n \rangle$ ,  $\mathcal{T}$ , and  $\mathcal{Q}$ , and choose  $k, n, m$  so  $f: V_n \rightarrow V_m$  and  $g: V_k \rightarrow V_n$ ; then  $f \circ g: V_k \rightarrow V_m$ . Construct  $T_k$ ,  $T_n$ , and  $T_m$  using  $\mathcal{T}$  and  $\langle V_n \rangle$ , and  $Q_k$ ,  $Q_n$ , and  $Q_m$  from those and  $\mathcal{Q}$ .

First we show  $\mathcal{P}$  is applicable to  $f \circ g$ . Pick any  $t_o \in T_m$ . Since  $\mathcal{P}$  is applicable to  $f$ ,  $t_{i_1} = t_o \circ f \in T_n$ . Similarly,  $\mathcal{P}$  is applicable to  $g$ , so  $t_{i_2} = t_{i_1} \circ g \in T_k$ . But the meaning of  $\mathcal{P}$  being applicable to  $f \circ g$  is just that (for each such  $t_o$ )  $t_o \circ (f \circ g) \in T_k$ , which is equivalent to  $(t_o \circ f) \circ g \in T_k$  which we just established.



Next we need to show (3), from which (4) and (5) will follow trivially. Recalling the earlier definition of  $\text{DualPatMap}$ , we first ensure the composition exists:

$$\text{DualPatMap}_{\mathcal{P}}(f): Q_m^* \rightarrow Q_n^*$$

$$\text{DualPatMap}_{\mathcal{P}}(g): Q_n^* \rightarrow Q_k^*$$

therefore

$$\text{DualPatMap}_{\mathcal{P}}(g) \circ \text{DualPatMap}_{\mathcal{P}}(f): Q_m^* \rightarrow Q_k^*$$

Now we'll evaluate each side of (3), applied to the same expression for  $t$  used in the definition of  $\text{DualPatMap}$ , and recalling that each function (or composition of functions) which is an element of some  $T_i$  is also a basis element of the corresponding  $Q_i^*$ . On the left,

$$\text{DualPatMap}_{\mathcal{P}}(f \circ g) \left( \sum_{t_o \in T_m} a_{t_o} t_o \right) = \sum_{t_o \in T_m} a_{t_o} (t_o \circ (f \circ g))$$

On the right,

$$\begin{aligned} & (\text{DualPatMap}_{\mathcal{P}}(g) \circ \text{DualPatMap}_{\mathcal{P}}(f)) \left( \sum_{t_o \in T_m} a_{t_o} t_o \right) \\ &= \text{DualPatMap}_{\mathcal{P}}(g) \left( \text{DualPatMap}_{\mathcal{P}}(f) \left( \sum_{t_o \in T_m} a_{t_o} t_o \right) \right) \\ &= \text{DualPatMap}_{\mathcal{P}}(g) \left( \sum_{t_o \in T_m} a_{t_o} (t_o \circ f) \right) \\ &= \sum_{t_o \in T_m} a_{t_o} \text{DualPatMap}_{\mathcal{P}}(g) (t_o \circ f) && \text{(by linearity)} \\ &= \sum_{t_o \in T_m} a_{t_o} ((t_o \circ f) \circ g) \end{aligned}$$

So the two sides are equal when applied to any  $t$ , which establishes (3) and thus the theorem. ■

## 4.4 measured complexity functions

Now we define  $C_{M,\mathcal{P}}$ , our basic proposed form for a measured complexity formula, and prove it valid. This form takes as parameters a pattern basis  $\mathcal{P}$  and a submultiplicative matrix measure  $M$ . (Section 2.17 shows computations using this definition,

and compares its value for various choices of  $M$  and  $\mathcal{P}$ . We hope, but don't know, that an  $M$  as simple as  $M_{\text{nz}}$  or  $M_{\text{abs}}$  will be sufficient to make some  $\mathcal{P}$  useful.)

This definition and its “validity” (as a nonnegative, “subadditive” measure with respect to function composition) don't depend on a choice of circuit model (except for requiring that  $\mathcal{P}$  is “applicable” to the model, i.e. that  $\mathcal{P}$ 's  $\langle T_n \rangle$  can compose with the model's constructible functions). But whether  $\mathcal{P}$  is *useful* (as opposed to producing only a trivial lower bound (after scaling) on circuit complexity for any  $f$ ) is likely to depend on the circuit model, in the sense that for each substantially different circuit model, we're likely to need a specially tailored  $\mathcal{P}$  (and perhaps  $M$ ), which will be hard to discover (and none is yet known). But the proof of validity is trivial, and (given applicability) is independent of  $\mathcal{P}$ ,  $M$ , and the circuit model.

Following the discussion in the Summary, we first define an “unscaled” measure,  $C_{M,\mathcal{P}}$ , and show it to be nonnegative and “subadditive” for function composition. (This means it would lowerbound complexity for compositions if it did for their components, but says nothing about whether it lowerbounds complexity for components, i.e. for our circuit model's generating circuits.)

We then define a “scaled” version,  $C_{M,\mathcal{P},\lambda} = \lambda C_{M,\mathcal{P}}$  for constant  $\lambda > 0$ , and show (trivially) that *if* it lowerbounds all generating circuit costs (in a given circuit model), then it lowerbounds the actual circuit complexity of all functions constructible in that model.

As discussed in the Summary, two issues can come up when trying to properly scale a complexity measure: it might not recognize some generating circuits as “free” (where “free” means  $C_{M,\mathcal{P}}(g) = 0$ , or strictly speaking,  $C_{M,\mathcal{P}}(f_{\text{circ}}(g)) = 0$ ) which the model thinks should be free ( $C_{\text{cost}}(g) = 0$ ); and for the generating circuits of positive cost, the necessary scale factor (to make  $C_{M,\mathcal{P}}$  lowerbound them) might have no positive lower limit (since there are generating circuits for every circuit width in  $\mathbb{N}$ ). (And even when  $C_{M,\mathcal{P}}$  can be correctly scaled, the resulting bounds might be trivial.) So later subsections here will parallel the less formal discussion in the Summary (especially the part starting at Section 2.14) about what it would take to use these theorems and address those issues to get nontrivial results — first for the ideal case of a perfectly-behaved pattern basis, then for various “fallback” cases.

#### 4.4.1 unscaled version ( $C_{M,\mathcal{P}}$ )

##### **Definition ( $C_{M,\mathcal{P}}$ ):**

Given a pattern basis  $\mathcal{P}$  (which defines the value spaces  $\langle V_n \rangle = \mathcal{P}_{\langle V_n \rangle}$ ), a function  $f: V_n \rightarrow V_m$  to which  $\mathcal{P}$  is applicable, and a submultiplicative matrix measure  $M$  whose value (for any pattern matrix produced by  $\mathcal{P}$ ) is at least 1, we define the

(unscaled) measured complexity formula  $C_{M,\mathcal{P}}$  as the log of  $M$ 's value on the pattern matrix for  $f$ :

$$C_{M,\mathcal{P}}: \text{Maps}(V_n, V_m) \rightarrow \mathbb{R}^{\geq 0}$$

$$C_{M,\mathcal{P}}(f) = \log_2 M(\text{PM}_{\mathcal{P}}(f))$$

(Recall that in a typical  $\mathcal{P}$  for general boolean functions,  $V_n = \mathbb{F}_2^n$ .)

**Theorem 3**  *$C_{M,\mathcal{P}}$  is a nonnegative, “subadditive” measure (with respect to function composition), for functions to which  $\mathcal{P}$  is applicable.*

**Proof.**  $C_{M,\mathcal{P}}$  is nonnegative, since by our assumption on  $M$ ,

$$C_{M,\mathcal{P}}(f) = \log_2 M(\text{PM}_{\mathcal{P}}(f)) \geq \log_2 1 = 0$$

Since pattern matrices multiply when their functions are composed (Theorem 2),

$$\text{PM}_{\mathcal{P}}(f \circ g) = \text{PM}_{\mathcal{P}}(f) \text{PM}_{\mathcal{P}}(g)$$

Since  $M$  is submultiplicative, we can apply it to get

$$M(\text{PM}_{\mathcal{P}}(f \circ g)) \leq M(\text{PM}_{\mathcal{P}}(f)) M(\text{PM}_{\mathcal{P}}(g))$$

Taking logs,

$$\log_2 M(\text{PM}_{\mathcal{P}}(f \circ g)) \leq \log_2 M(\text{PM}_{\mathcal{P}}(f)) + \log_2 M(\text{PM}_{\mathcal{P}}(g))$$

or in other words

$$C_{M,\mathcal{P}}(f \circ g) \leq C_{M,\mathcal{P}}(f) + C_{M,\mathcal{P}}(g)$$

■

(It might be possible to define a “subadditive” matrix measure (at least for pattern matrices, which will presumably have more structure than general matrices) in some other way than as the log of a submultiplicative measure (perhaps analogously to entropy?); if so, that kind of measure could work just as well in this general framework, and would be worth exploring.)

#### 4.4.2 scaled version ( $C_{M,\mathcal{P},\lambda}$ )

Terminology:  $C_{M,\mathcal{P},\lambda}$  is just an abbreviation for  $\lambda C_{M,\mathcal{P}}$ , for any constant  $\lambda > 0$ . Sometimes we might slip and call  $C_{M,\mathcal{P},\lambda}$  a “*scaled complexity measure*” unconditionally, but it’s better to reserve that term for when it actually lowerbounds circuit complexity (in a given circuit model). But we’ll prove it’s enough to know it does that for the generating circuits — lowerbounding actual complexity for all constructible functions then follows by subadditivity.

In the following, we fix a circuit model (with a set  $\mathcal{G}$  of generating circuits, a set  $\mathcal{F}$  of constructible functions, a circuit cost function  $C_{\text{cost}}: \mathcal{G} \rightarrow \mathbb{R}^{\geq 0}$  obeying all rules in the definition of “circuit model” in Section 2.1.1, and a corresponding circuit complexity function  $C_{\min}: \mathcal{F} \rightarrow \mathbb{R}^{\geq 0}$ ), a pattern basis  $\mathcal{P}$  applicable to that model, a submultiplicative matrix measure  $M$  which meets the conditions in the definition of  $C_{M,\mathcal{P}}$  (so Theorem 3 shows  $C_{M,\mathcal{P}}$  is a “subadditive” measure for function composition), and a positive real constant  $\lambda$ . We then have:

**Theorem 4** *If  $C_{M,\mathcal{P},\lambda}$  lowerbounds circuit cost for all generating circuits (in  $\mathcal{G}$ ), then it lowerbounds circuit complexity for all constructible functions (in  $\mathcal{F}$ ). That is,*

$$\forall_{g \in \mathcal{G}}: C_{M,\mathcal{P},\lambda}(g) \leq C_{\text{cost}}(g) \implies \forall_{f \in \mathcal{F}}: C_{M,\mathcal{P},\lambda}(f) \leq C_{\min}(f)$$

(As usual, by  $C_{M,\mathcal{P},\lambda}(g)$  for a circuit  $g$ , we mean  $C_{M,\mathcal{P},\lambda}(f_{\text{circ}}(g))$ , where  $f_{\text{circ}}(g)$  denotes the function computed by  $g$ .)

**Proof.** For any  $f \in \mathcal{F}$ ,  $C_{\min}(f)$  is the minimum cost of any circuit which computes  $f$ ; as discussed where “circuit models” were defined (Section 2.1.1), it’s also the minimum cost over compositions of generating circuits  $g_i \in \mathcal{G}$  which compute  $f$ , i.e. for which  $f = g_1 \circ \dots \circ g_k$ . This means that for at least one such composition  $\langle g_i \rangle$ ,  $C_{\min}(f) = \sum_i C_{\text{cost}}(g_i)$ .

But we also have, by assumption,  $C_{\text{cost}}(g_i) \geq C_{M,\mathcal{P},\lambda}(g_i)$  for each  $g_i$ . Summing those and combining,

$$C_{\min}(f) = \sum_i C_{\text{cost}}(g_i) \geq \sum_i C_{M,\mathcal{P},\lambda}(g_i)$$

On the other hand, since (by Theorem 3)  $C_{M,\mathcal{P}}$  is “subadditive” for function composition,

$$\lambda C_{M,\mathcal{P}}(f) = \lambda C_{M,\mathcal{P}}(g_1 \circ \dots \circ g_k) \leq \sum_i \lambda C_{M,\mathcal{P}}(g_i)$$

Combining, we have

$$C_{M,\mathcal{P},\lambda}(f) \leq C_{\min}(f)$$

Since this holds for every  $f \in \mathcal{F}$ , we prove the theorem. ■

## 4.5 “conventional” structure of generating circuits

For the upcoming theorems, it will be convenient to assume that all generating circuits  $g \in \mathcal{G}$  have a “conventional” structure (which we’ll define here, and which is illustrated in Section 2.1.2). We’ll then assume this for the rest of Section 4.

(This assumption is not quite “without loss of generality”, but I don’t know of any *interesting* circuit models which can’t easily be modified to satisfy it. Note that we’re not requiring the model to define *all* generating circuits which have the following structures — only that whichever ones it *does* define fit into this scheme.)

For any generating circuit  $g \in \mathcal{G}$  (or any other circuit or gate  $g$  used below), we’ll denote its number of inputs by  $n_g \geq 0$  and its number of outputs by  $m_g \geq 0$ . Such a  $g$  computes a function  $f_{\text{circ}}(g): V_{n_g} \rightarrow V_{m_g}$ , for value spaces  $V_n$  suitable for a pattern basis applicable to this circuit model (typically,  $V_n = \mathbb{F}_2^n$ ). If we say below that  $n_g$  and  $m_g$  are “small”, in theory we just mean “constant” (that is, limited by some constant of the model, not specific to  $g$ ), but in practice we really mean *small*, since values up to 3 are enough for a typical application.

So, we now assume that all generating circuits  $g \in \mathcal{G}$  are one of the following:

- a **permutation of wires** (so  $m_g = n_g$ )
- a **core circuit** (from a finite set  $\mathcal{G}_{\text{core}}$ , so  $n_g$  and  $m_g$  are “small”), which is one of:
  - a **primitive gate** (which might have any circuit-function  $f_{\text{circ}}(g): V_{n_g} \rightarrow V_{m_g}$ , for value spaces  $V_n$  suitable for a pattern basis)
  - a **trivial wire operation**, which means one of:
    - \* splitting a wire (replacing it with two wires, each carrying the same value as the original, so  $(n_g, m_g) = (1, 2)$ ; any uninvolved wires are not considered part of this  $g$ , but will be added as “adjoined wires” by the “widened core circuit” case below)
    - \* discarding a wire (so  $(n_g, m_g) = (1, 0)$ )
    - \* creating a wire (whose value is some constant; this has  $(n_g, m_g) = (0, 1)$ )
- a **widened core circuit**, which consists of a core circuit  $g'$  with  $k \geq 1$  non-interacting wires adjoined in parallel, by the circuit-widening “adjoin wires operation”  $\text{aw}_k$  (so  $g = \text{aw}_k(g')$ , giving  $(n_g, m_g) = (n_{g'} + k, m_{g'} + k)$ ).<sup>36</sup>

---

<sup>36</sup>  $\text{aw}_k$  itself is defined more generally, for  $k \geq 0$  and any circuit  $g$  or function  $f$ , in a way that

(If we wished, we could simplify this (WLOG) by considering the trivial wire operations to be primitive gates, and by permitting  $k = 0$  in the widened core circuits. But the above description clarifies our discussion; to simplify our reasoning, we can in general talk about core circuits rather than primitive gates. If we ever need the **empty gate** (whose  $(n_g, m_g) = (0, 0)$ ), we can count it as a primitive gate (so it’s included in  $\mathcal{G}_{\text{core}}$ ).)

This structural assumption says nothing about *costs* of generating circuits — we’ll cover that in the subsequent subsections. We’re also not assuming anything here about *which* circuits with the above structure are actually present in  $\mathcal{G}$ . But later subsections may make such assumptions locally; in particular, if they say “assume wire permutations are free” (meaning they satisfy  $C_{\text{cost}}(g) = 0$ ), this should also be taken as implying that all of them exist in  $\mathcal{G}$ . Similarly, if they upperbound  $C_{\text{cost}}(g)$  for some circuit  $g$  which has this structure, this should be understood as implying  $g \in \mathcal{G}$ , unless stated otherwise.

## 4.6 simple case — pattern basis with ideal behavior

A typical circuit model for general boolean computation defines all wire permutations and trivial wire operations as free ( $C_{\text{cost}}(g) = 0$ ), and also counts non-interacting wires as free (when they’re adjoined in parallel to other circuits). An ideal complexity formula would do the same.

Since we hope this ideal behavior can be achieved, and since its consequences are simplest to analyze, it makes sense to spell them out formally. (The later “fallback” sections discuss what to do if we can’t achieve this behavior.) To summarize the informal analysis in and around Section 2.14, which we follow here, our basic conclusion is just that measured complexity can then be scaled (by a small factor) to become a lower bound on actual circuit complexity.

First, let’s define this “ideal behavior” more formally. About  $C_{\text{cost}}$  itself:

- For our formal conclusions here, it doesn’t matter whether the trivial wire operations are free.<sup>37</sup>

---

commutes with  $f_{\text{circ}}$ . The inputs and outputs of a circuit or function are ordered; we define  $\text{aw}_k$  as extending both sequences at the end. It then “commutes with composition” as well:  $\text{aw}_k(g) \circ \text{aw}_k(h) = \text{aw}_k(g \circ h)$  (whenever  $g \circ h$  exists; to be technical, for circuits (in some models) this might be “equivalence” (i.e. equality of  $f_{\text{circ}}$  values) rather than equality of the circuits themselves).

<sup>37</sup> If they’re not, it does change the meaning of whatever lower bounds on  $C_{\text{min}}$  we come up with — but only linearly (as shown in Section 2.14). This is discussed further in the “fallback” section.

- But we do assume  $C_{\text{cost}}$  considers at least the wire permutations free.<sup>38</sup>
- Finally, we assume that adjoining wires (when permitted) doesn't change the cost defined for any generating circuit.<sup>39</sup>

About  $C_{M,\mathcal{P}}$ , what we need is basically the same, except we should also ensure it's at least as generous as  $C_{\text{cost}}$  about considering other generating circuits free (beyond just wire permutations). Formally, we need:

- $C_{M,\mathcal{P}}$  understands that (at least) the same generating circuits should be free:

$$\forall_{g \in \mathcal{G}}: C_{\text{cost}}(g) = 0 \implies C_{M,\mathcal{P}}(g) = 0 \quad (6)$$

(Note that  $C_{M,\mathcal{P}}$ 's subadditivity then implies the same thing for all constructible functions:  $\forall_{f \in \mathcal{F}}: C_{\min}(f) = 0 \implies C_{M,\mathcal{P}}(f) = 0$ .<sup>40</sup> Ideally this implication would go in both directions, but we can ignore that here, since we'll later assume separately that  $C_{M,\mathcal{P}}$  sees high (unscaled) measured complexity in the functions we want it to.)

- Adjoining non-interacting wires should not increase (unscaled) measured complexity, at least for generating circuits.<sup>41</sup> Using our assumption about their structure, and recalling that  $\text{aw}_k(g)$  denotes the circuit made by adjoining  $k$  non-interacting wires to  $g$  (and that  $\mathcal{G}$  is not required to contain  $\text{aw}_k(g)$ , though typically it would), we can state this as:

$$\forall_{g \in \mathcal{G}, k \geq 0}: \text{aw}_k(g) \in \mathcal{G} \implies C_{M,\mathcal{P}}(\text{aw}_k(g)) \leq C_{M,\mathcal{P}}(g) \quad (7)$$

(Unlike with (6), this *doesn't* necessarily extend to all constructible functions. Fortunately we don't need that for our conclusions here.)

---

<sup>38</sup> We use that explicitly, but it's important to note that we use it implicitly too, to justify the reasonableness of our structural assumption about generating circuits — otherwise we'd need to consider “adjoining a wire on either end of the sequences of inputs and outputs” (not just on one end, like  $\text{aw}_k$  does now), to get the generating circuits needed in a typical circuit model.

<sup>39</sup> It would be possible to construct a circuit model in which adjoining a wire sometimes *decreased* a generating circuit's cost, and perhaps even to imagine this being reasonable for some applications, but this would complicate our analysis here, so it seems better to define it as “non-ideal behavior”. Thus we say “doesn't change” rather than “doesn't increase”.

<sup>40</sup> This follows from observing that the zero value of  $C_{\min}(f)$ , defined as a minimum over compositions of generating circuits  $\langle g_i \rangle$ , must be achieved by some composition of free  $g_i$ .

<sup>41</sup> We don't expect this to *decrease* it either (i.e. we expect to have equality here), but we don't need that assumption.

It's now straightforward to formalize the proof outlined in Section 2.15:

**Lemma 5** *Given our assumption of “ideal behavior”, there exists a small positive  $\lambda$  (depending only on  $C_{M,\mathcal{P}}$  and  $C_{\text{cost}}$ , both restricted to  $\mathcal{G}_{\text{core}}$ , and with  $1/\lambda$  also small) which will scale  $C_{M,\mathcal{P}}$  to lowerbound  $C_{\text{cost}}$  for all  $g \in \mathcal{G}$ ; that is, we'll have*

$$\forall_{g \in \mathcal{G}}: \lambda C_{M,\mathcal{P}}(g) \leq C_{\text{cost}}(g)$$

(Note that this will give us the condition we need to apply Theorem 4.)

**Proof.** Whenever  $C_{M,\mathcal{P}}(g) = 0$ , the inequality  $\lambda C_{M,\mathcal{P}}(g) \leq C_{\text{cost}}(g)$  is trivial, so we can ignore such  $g$  in determining  $\lambda$ .<sup>42</sup>

To cover the rest of  $\mathcal{G}$ , just set  $\lambda$  to the minimum, over the other  $g \in \mathcal{G}_{\text{core}}$  (i.e. over the core circuits which have positive measured complexity), of  $C_{\text{cost}}(g)/C_{M,\mathcal{P}}(g)$ :

$$\lambda = \min_{g \in \mathcal{G}_{\text{core}}, C_{M,\mathcal{P}}(g) > 0} C_{\text{cost}}(g)/C_{M,\mathcal{P}}(g)$$

This is positive, since “ideal behavior” includes the condition  $C_{\text{cost}}(g) = 0 \implies C_{M,\mathcal{P}}(g) = 0$  (so no term can be 0), and we minimized over a finite set.

The only  $g$  this didn't explicitly cover are the wire permutations, but they're “free” (0-valued) in both measures; and the widened core circuits, but the ratio we're minimizing can't decrease (compared to its value for the core circuit  $g'$  they're based on, which is already in the minimum), so including those  $g$  would not change the minimum.<sup>43</sup>

(Finally, we can fairly describe all values of  $C_{\text{cost}}(g)$  or  $C_{M,\mathcal{P}}(g)$  used in the ratio we minimized as “small”, since they all come from measures of core circuits. Thus both  $\lambda$  and its inverse are “small”.<sup>44</sup>) ■

Using the  $\lambda$  from Lemma 5 to define  $C_{M,\mathcal{P},\lambda}$ , we now have just what we need for applying Theorem 4, to get:

**Proposition 6** *The  $C_{M,\mathcal{P},\lambda}$  just defined lowerbounds  $C_{\min}$  for all constructible functions:*

$$\forall_{f \in \mathcal{F}}: C_{M,\mathcal{P},\lambda}(f) \leq C_{\min}(f)$$

---

<sup>42</sup> If this means we ignore *all*  $g$ , we can pick any  $\lambda > 0$ . This implies  $C_{M,\mathcal{P}}(f) = 0$  for every constructible function  $f$  (since  $C_{M,\mathcal{P}}$  is subadditive and  $\mathcal{G}$  generates all constructible functions).

<sup>43</sup> They have the same  $C_{\text{cost}}(g)$ , and the same or lower  $C_{M,\mathcal{P}}(g)$ , as  $g'$ . If  $C_{M,\mathcal{P}}(g)$  decreases to 0, we were already ignoring that  $g$ .

<sup>44</sup> As stated before, formally “small” just means “constant”, but in practice it's an accurate informal term. In this case, we expect both those measures to range from about 1 to at most  $\max_{g \in \mathcal{G}_{\text{core}}} m_g 2^{n_g}$ .



Furthermore,  $1/\lambda$  is “small” (in the sense described earlier), so for any  $f$  with “high”  $C_{M,\mathcal{P}}(f)$ , we get a significant (proportional) lowerbound on  $C_{\min}(f)$ .

**Proof.** Apply Theorem 4 (and recall that  $C_{M,\mathcal{P},\lambda}(f) = \lambda C_{M,\mathcal{P}}(f)$ ). ■

To complete our more formal parallel of the claims in and around Section 2.14, all that remains is to point out that if we can find an “ideally behaved”  $C_{M,\mathcal{P}}(f)$  which has a “high” value for a lot of  $f$ , or for any interesting  $f$ , we can apply Proposition 6 to get significant lower bounds on  $C_{\min}(f)$  for those  $f$ .

(We didn’t state the above result “asymptotically”, since in practice it needn’t be applied that way, but it could be — what we showed is essentially  $C_{\min}(f) \geq \Omega(C_{M,\mathcal{P}}(f))$ , so (for example) for any explicit function family  $\langle f_n \rangle$  with  $f_n$  having input size  $n$ , if  $C_{M,\mathcal{P}}(f)$  is superlinear or superpolynomial in  $n$ , then  $C_{\min}(f)$  has the same property.)

## 4.7 fallbacks (for worse behavior)

It’s possible we could fail to find a pattern basis  $\mathcal{P}$  (or a suitable  $M$ ) which gives  $C_{M,\mathcal{P}}$  “ideal behavior” in the sense of Section 4.6, but could still find one with good enough behavior to get interesting conclusions. So in this subsection we outline various weaker but still interesting bounds we could derive from less ideal behavior of  $C_{M,\mathcal{P}}$ . (In the worst case these would be polynomially worse bounds, in the sense that (what we guess to be) the actual complexity of some explicit function family  $\langle f_n \rangle$  might be some polynomial (of arbitrary degree, but independent of  $n$ ) applied to the best lower bound we could prove using these “fallback” methods; but that doesn’t rule out applying them to a proof of  $\text{NP} \not\subseteq \text{P/poly}$ , which is a very weak bound compared to what we expect to be true.<sup>45</sup> Note that these methods are not relevant to looking for an ideal pattern basis — the sufficiently optimistic reader can safely skip this subsection.)

We’ll assume our circuit model is “fully conventional” — that is,  $\mathcal{G}$  has the conventional structure described in Section 4.5, and includes a usual set of operations (core circuits) sufficient for general boolean computation, with the usual division regarding what is free (wire permutations and trivial wire operations, and adjoining wires to any circuit) vs. not free (everything else). (This implies, but is stronger than, the set of conditions on  $C_{\text{cost}}$  listed in Section 4.6.)

---

<sup>45</sup> For example, the Exponential Time Hypothesis [IP01] is widely considered likely; roughly, it asserts that some problems in  $\text{NP}$  require exponential time. As far as I know, a similar guess about required non-uniform circuit complexity for  $\text{NP}$  is also widely held.

Given the above, there are three progressively-worse kinds of “non-ideal” behavior we might face in  $C_{M,\mathcal{P}}$  (aside from its failing to see *any*  $f$  as having high measured complexity):

1. there are  $g \in \mathcal{G}$  which are free in  $C_{\text{cost}}$ , but not free according to  $C_{M,\mathcal{P}}$  — but these are all trivial wire operations (not wire permutations);
2. even some wire permutations are not free in  $C_{M,\mathcal{P}}$ ;
3. or worst of all,  $C_{M,\mathcal{P}}$  hasn’t solved the “adjoined-wire issue” — that is, for at least some  $g \in \mathcal{G}$  and some  $k > 0$ ,

$$C_{M,\mathcal{P}}(\text{aw}_k(g)) > C_{M,\mathcal{P}}(g).$$

We won’t always be able to recover from these problems (except for problem 1), but the following subsections outline some ways we can try.<sup>46</sup> (For brevity, each subsection frequently uses results from the prior ones (though we try to state each instance of this explicitly). Each subsection assumes the earlier-listed problems may be occurring at the same time, but not the later-listed problems; this lets us solve all three problems at once (presuming each can be solved on its own, when we get to it), by “working from the end”, and combining the bounds-weakenings we get from each step.)

#### 4.7.1 not all trivial wire operations are free

Problem 1 has already been covered in Section 2.14 (with an accurate if sketchy proof), but we spell it out more formally here. (The proof here is exactly the same — just explained much more carefully. Unfortunately that makes it much longer, so the referenced discussion should be read first.)

Suppose some of the “trivial wire operations” (defined in that section, and again with more detail in Section 4.5 — namely, splitting or copying, discarding, or creating a wire) have positive measured complexity.

Informally, we might think of these operations as applicable to any circuit width  $n$  — for example, a generating circuit for splitting the  $i$ th wire (out of  $n$ ) might produce a new  $n + 1$ st output wire (whose value equals that of the  $i$ th input wire), but also pass all  $n$  input wires unchanged to same-numbered outputs (so its  $(n_g, m_g)$

---

<sup>46</sup> These include relations between minimum cost in different circuit models which are straightforward and seem likely to be well-known, but I give full proofs here since I have no references for them.

would be  $(n, n + 1)$ ). Then if  $C_{M,\mathcal{P}}(g)$  was no lower for this  $g$  than for an arbitrary function of the same size, this single  $g$  might have as much measured complexity as the whole function  $f$  (computed by some circuit that includes  $g$ ) whose complexity we’re trying to bound (or even more, if  $f$  has fewer than  $n$  inputs) — a fatal problem.

But fortunately, we don’t have to think of these operations that way. As long as the “adjoined-wire issue” is solved, and wire permutations are free (both of which we’re still assuming here), we can “factor” that operation into a *small* wire-splitting circuit (with  $(n_g, m_g) = (1, 2)$ , regardless of  $n$ ), with  $n - 1$  extra wires adjoined to it in parallel (using  $\text{aw}_{n-1}$ , for no increase in measured complexity), composed with wire permutations before and after. (And we can do this for *all* trivial wire operations, with the necessary core circuit having that size or smaller.)

Then our assumptions guarantee the factored circuit’s measured complexity is no worse than the worst case of  $C_{M,\mathcal{P}}(f)$  over all  $f$  of size  $(n, m) = (1, 2)$  or smaller, which is constant (since only a small finite number of functions have that size).

So we needn’t worry about trivial wire operations of *arbitrary* cost — only of *constant* cost.<sup>47</sup>

We then observe that in any circuit (which computes  $f: \mathbb{F}_2^n \rightarrow \mathbb{F}_2^m$ ) made of  $X$  primitive gates, with each gate  $g_i$  having  $n_{g_i}$  inputs and  $m_{g_i}$  outputs (note that these gates are core circuits — the generating circuits that contain them would usually be wider), that no more than  $m + n + \sum_i (n_{g_i} + m_{g_i})$  trivial wire operations are needed — one for each input or output of any gate, or of the whole function. (Each gate input or whole-circuit output might need to copy a wire or create a constant, and each gate output or whole-circuit input might at some stage need to be discarded.)

(Of course we *could* create circuits which use more trivial operations than that, but no circuit *needs* to (to compute whatever function it computes) — it’s always possible to revise any circuit to remove the excessive operations, without increasing its cost (by either measure), or changing anything about the gates themselves (their order or “what they compute”), or changing what the whole circuit computes.<sup>48</sup>)

To take advantage of this observation, we’ll create a modified circuit model, with the same  $\mathcal{G}$  but a new cost function  $C'_{\text{cost}}$ , which is based on  $C_{\text{cost}}$  but differs for the trivial wire operations. (We’ll then become able to lowerbound the modified  $C'_{\text{cost}}$ ,

---

<sup>47</sup> For brevity, I’m sometimes saying “cost” when I ought to say “measured complexity” — though this shorthand could also be considered a preview of a later step in the argument, where we’ll introduce a “modified cost” which better matches the measured complexity.

<sup>48</sup> I’m not spelling out the details completely formally here, since these facts are presumably both obvious and well-known. But I’ll point out that in the revised circuit, every gate input (or whole-circuit output) receives an identical “signal” as in the original circuit, where by “signal” we mean the function of the  $n$  whole-circuit inputs which gives the value carried by a specified wire.

and transfer this back to a weaker bound on the original  $C_{\text{cost}}$ .) Specifically,  $C'_{\text{cost}}$  should see the trivial wire operations as having a “more realistic” cost — positive rather than 0 — but be otherwise unchanged from  $C_{\text{cost}}$ .<sup>49</sup>

We can then define  $C'_{\min}: \mathcal{F} \rightarrow \mathbb{R}^{\geq 0}$  from  $C'_{\text{cost}}$  in the same way as  $C_{\min}$  was defined from  $C_{\text{cost}}$  — each is the minimum possible cost (using their respective cost functions) of any circuit (composition of functions computed by  $g_i \in \mathcal{G}$ , for a given sequence  $\langle g_i \rangle$ ) which computes their argument function  $f$ .

We now have two parallel goals — use Lemma 5 and Theorem 4 to show that  $C_{M,\mathcal{P}}$  can be scaled to lowerbound the *modified*  $C'_{\min}$  (even though it’s unable to lowerbound the original  $C_{\min}$ ); and use our observation about how many trivial wire operations can be needed (in a circuit with  $X$  primitive gates) to transfer a lower bound from  $C'_{\min}$  to  $C_{\min}$  (for any  $f$ ). (We’ll do the second one first.)

To relate  $C'_{\min}$  and  $C_{\min}$ , consider a minimum circuit for  $f$  under  $C_{\text{cost}}$  (as  $C_{\min}$  does). It will express  $f$  as a composition of some sequence  $\langle g_i \rangle$  of generating circuits, which (say) contains  $X$  primitive gates, which we’ll call  $\langle G_j \rangle$ . (We name the primitive gates using capital  $G$  to avoid confusion, since in general each  $G_j$  is distinct from any  $g_i$  — instead, it will be the core circuit of some  $g_i$ .)

Then by definition of  $C_{\min}$ ,

$$C_{\min}(f) = \sum_i C_{\text{cost}}(g_i)$$

but due to our various assumptions (which make everything other than primitive gates free in  $C_{\text{cost}}$ ), we can reduce this to simply the cost from the primitive gates alone:

$$C_{\min}(f) = \sum_j C_{\text{cost}}(G_j)$$

Now,  $C'_{\min}$  has its own ideas about cost, so it might think some other circuit is better, but it can’t do *worse* than if it used that same circuit, so (recalling that it agrees with  $C_{\text{cost}}$  about gate costs) we have:

$$C'_{\min}(f) \leq \sum_j C_{\text{cost}}(G_j) + \sum_k C'_{\text{cost}}(W_k)$$

where  $\langle W_k \rangle$  are the trivial wire operations used in the same circuit  $\langle g_i \rangle$ .

---

<sup>49</sup> We’ll get a *correct* bound even if those positive costs are arbitrarily chosen (as long as adjoining wires doesn’t change  $C'_{\text{cost}}$ , as we already assumed about  $C_{\text{cost}}$ ). To optimize the implicit constant factor, the modified costs should be proportional to the corresponding measured complexities.

This sequence  $\langle W_k \rangle$  might have any length (compared to the number of primitive gates  $X$ ), but without loss of generality we can ask our earlier use of  $C_{\min}(f)$  to pick a  $\langle g_i \rangle$  which minimizes that length (while fixing the sequence of gates  $\langle G_j \rangle$  and their “input signals”), since doing so doesn’t change the minimum cost it finds. But we earlier showed, about the minimum length  $Y$  which  $\langle W_k \rangle$  can be made to have, that

$$Y \leq m + n + \sum_j (n_{G_j} + m_{G_j})$$

(where as usual  $f: \mathbb{F}_2^n \rightarrow \mathbb{F}_2^m$ , and gate  $G_j$  has size  $(n_{G_j}, m_{G_j})$ ).

Noting that  $\mathcal{G}_{\text{core}}$  is finite (and contains every  $G_j$ ), set  $N$  and  $M$  to the (constant) maximum possible values of  $n_{G_j}$  and  $m_{G_j}$  respectively; then

$$Y \leq m + n + X(N + M)$$

This is now enough to let us transfer a bound. First we note that finiteness of sets (and our other assumptions) give us constants for a maximum possible  $C'_{\text{cost}}(W_k)$ , and for a (positive) minimum possible  $C_{\text{cost}}(G_j)$ ; we’ll suppress these constants using the notations  $O()$  and  $\Omega()$ . Combining conclusions from above, we then have:

$$\begin{aligned} C'_{\min}(f) &\leq \sum_j C_{\text{cost}}(G_j) + \sum_k C'_{\text{cost}}(W_k) \\ &\leq C_{\min}(f) + YO(1) \\ &\leq C_{\min}(f) + (m + n + X(N + M))O(1) \end{aligned}$$

But also

$$C_{\min}(f) = \sum_j C_{\text{cost}}(G_j) \geq \Omega(X)$$

or in other words

$$X \leq O(C_{\min}(f))$$

which lets us continue the above as:

$$\begin{aligned} C'_{\min}(f) &\leq C_{\min}(f) + (m + n + X(N + M))O(1) \\ &\leq C_{\min}(f) + (m + n)O(1) + O(C_{\min}(f)) \\ &\leq O(C_{\min}(f)) + O(m + n) \\ &\leq O(C_{\min}(f) + \alpha(m + n)) \end{aligned}$$

(where  $\alpha > 0$  can be chosen arbitrarily for our later convenience).

Turning this around,

$$C_{\min}(f) + \alpha(m + n) \geq \Omega(C'_{\min}(f))$$

and thus (for some  $\beta > 0$ , controllable by choosing  $\alpha$ ),

$$C_{\min}(f) \geq \Omega(C'_{\min}(f) - \beta(m + n))$$

which will let us transfer nontrivial (i.e. superlinear in  $n$  or better) lower bounds from  $C'_{\min}$  to  $C_{\min}$  as promised, in an essentially linear way.

Now we can turn to the other parallel goal we had earlier — using Lemma 5 and Theorem 4 to get a lower bound on  $C'_{\min}(f)$  in terms of  $C_{M,\mathcal{P}}(f)$ . But this is easy, since the only way  $C_{M,\mathcal{P}}$  was “non-ideal” was that it considered some core circuits non-free which  $C_{\text{cost}}$  considered free — but  $C'_{\text{cost}}$  considers them non-free too, and that was allowed by our requirements for the cost function when we defined “ideal behavior” — so *relative to*  $C'_{\text{cost}}$ ,  $C_{M,\mathcal{P}}(f)$  has “ideal behavior”. That means we can use Lemma 5 exactly as before, to find  $\lambda$  such that  $\lambda C_{M,\mathcal{P}}(g)$  lowerbounds  $C'_{\text{cost}}(g)$ , and then use Theorem 4 to conclude that  $\lambda C_{M,\mathcal{P}}(f)$  lowerbounds  $C'_{\min}(f)$  (for all  $f \in \mathcal{F}$ ).

Putting it all together: just as in the fully “ideal” situation, we find that for any  $f$  for which  $C_{M,\mathcal{P}}$  gives us a “high” (e.g. superlinear in  $n$ ) value, we get a nontrivial lower bound on  $C_{\min}(f)$  as well.

#### 4.7.2 even some wire permutations are not free

Problem 2 may or may not be practical to handle, depending on the details, so we’ll only cover one straightforward possibility here, which seems plausible in practice — namely, all we’ve directly proven about  $C_{M,\mathcal{P}}$  (concerning wire permutations) is that we can swap any two adjacent wires (out of the  $n$  wires entering any given circuit stage), at a cost (or more precisely, at a provable upper bound on measured complexity  $C_{M,\mathcal{P}}$ ) which is no higher than a fixed polynomial  $p$  of  $n$ . (By “two adjacent wires”, we mean a pair of coordinates indexed by  $i$  and  $i + 1$  in  $V_n = \mathbb{F}_2^n$ .<sup>50</sup>)

---

<sup>50</sup> If we wanted to formalize this further, we’d define  $f_{\text{swap}(i,j,n)}: \mathbb{F}_2^n \rightarrow \mathbb{F}_2^n$  as the function which swaps coordinates  $i$  and  $j$  in any vector in  $\mathbb{F}_2^n$ , and assume we’ve proven  $C_{M,\mathcal{P}}(f_{\text{swap}(i,i+1,n)}) \leq p(n)$ . Note that unlike in problem 1, we can’t reduce this  $p(n)$  to a constant, since the way we did that depended on the wire permutations being free in  $C_{M,\mathcal{P}}$  (as they are in  $C_{\text{cost}}$ ). (If the wires being swapped have low indices  $i$  compared to  $n$ , we might “de-adjoin wires” to reduce  $p(n)$  to  $p(i)$ , but this is not useful in general, since we need to be able to handle any  $i$ .)

First, note that the limitation of our knowledge about  $C_{M,\mathcal{P}}$  to certain permutations is only apparent. Since any permutation is a composition of  $O(n^2)$  swaps of adjacent pairs,  $C_{M,\mathcal{P}}$ 's “subadditivity” proves it's upperbounded by some other polynomial  $p'(n)$  (of degree 2 higher than  $p$ ) for an arbitrary permutation of wires.<sup>51</sup> That is, for any wire permutation  $f_0$  on  $\mathbb{F}_2^n$ ,  $C_{M,\mathcal{P}}(f_0) \leq p'(n)$ .

At this point we proceed similarly to problem 1, but with two changes. One is that our modified circuit cost  $C'_{\text{cost}}$  (in which permuting  $n$  wires costs  $p'(n)$ , rather than 0 as in  $C_{\text{cost}}$ ) will produce a  $C'_{\text{min}}$  polynomially higher than  $C_{\text{min}}$  (rather than just linearly higher). This will result in a weakening of any lower bound we transfer from  $C'_{\text{min}}$  to  $C_{\text{min}}$ , as described below. (For simplicity in expressing that transfer, we'll also ignore functions which require *only* wire permutations to compute. That way we won't need to do the sort of thing we did to handle the related issue in problem 1 (namely, bounding  $C_{\text{min}} + n + m$  rather than  $C_{\text{min}}$ ) to prevent pure wire permutations from being a counterexample, as they otherwise would be, due to their  $C'_{\text{min}}$  being arbitrarily high (as  $n$  grows), while their  $C_{\text{min}}$  remains 0.)

The other change is that  $C_{M,\mathcal{P}}$  is no longer “ideal” relative to *any* cost function (since that requires wire permutations being free in both  $C_{M,\mathcal{P}}$  and the cost function), so Lemma 5 will no longer be capable of finding the scaling factor  $\lambda$  we need. Fortunately, a trivial generalization will be enough:

**Lemma 7** *If, aside from wire permutations,  $C_{M,\mathcal{P}}$  has “ideal behavior” relative to  $C'_{\text{cost}}$  — but for wire permutations themselves, either  $C_{M,\mathcal{P}}$  is 0, or  $C'_{\text{cost}}$  is positive and  $C_{M,\mathcal{P}}/C'_{\text{cost}}$  is upperbounded — then there exists a  $\lambda$  with the same properties as in Lemma 5 (but for  $C'_{\text{cost}}$  rather than  $C_{\text{cost}}$ ).*

**Proof.** Same as for Lemma 5, except the minimum of  $C'_{\text{cost}}/C_{M,\mathcal{P}}$  (used to determine  $\lambda$ ) should also include the wire permutations with positive  $C_{M,\mathcal{P}}$ , whose contribution has a positive lower bound (since its inverse is upperbounded, by assumption). ■

To apply Lemma 7 here, just note that we're assuming arbitrary wire permutations of width  $n$  have  $C_{M,\mathcal{P}}$  at most  $p'(n)$ , but we've defined their  $C'_{\text{cost}}$  as exactly  $p'(n)$ .

Otherwise the situation is the same as in problem 1, except for its weaker result — Lemma 7 will find the  $\lambda$  which lets Theorem 4 show  $C_{M,\mathcal{P},\lambda}$  lowerbounds  $C'_{\text{min}}$ ; if (for

---

<sup>51</sup> Proof: consider the Cayley graph of permutations of  $n$  wires in which all swaps are (directed) edges; label all nodes and edges by their value of  $C_{M,\mathcal{P}}$  (assumed at most  $p(n)$  for edges, though not necessarily equal for one edge's two directions). The subadditivity of  $C_{M,\mathcal{P}}$  implies a “triangle inequality” which upperbounds (by  $p(n)$ ) the absolute difference between  $C_{M,\mathcal{P}}$  of adjacent nodes; but the graph is connected, with diameter  $O(n^2)$ , and  $C_{M,\mathcal{P}}$  of its “identity node” is 0.

some function family  $\langle f_n \rangle$  of increasing width  $n$ ) that lower bound is some polynomial in  $n$  of sufficiently high degree, we now show how to get a lower-degree but still polynomial lower bound on  $C_{\min}(f_n)$ . (And a superpolynomial lower bound on  $C'_{\min}$  would give us a superpolynomial lower bound on  $C_{\min}$  as well.)

Specifically, for that lower-bound-transfer to follow, we must show:

**Proposition 8** *If permuting  $k$  wires costs at most  $p'(k)$  in  $C'_{\text{cost}}$  (for some polynomial  $p'$ ), then there is a polynomial  $q$  such that, for all  $f \in \mathcal{F}$  (with  $f: \mathbb{F}_2^n \rightarrow \mathbb{F}_2^m$  for any  $n$  and  $m$ ) which are not just wire permutations,  $C'_{\min}(f) \leq q(C_{\min}(f))$ . Furthermore,  $q$  is independent of  $n$  and  $m$ , and  $\deg(q) = O(\deg(p'))$ .*

**Proof.** In a circuit composed from  $Y$  wire permutations and  $X$  other generating circuits, adjacent permutations could be replaced by a single one, so minimizing their number gives  $Y \leq X + 1 = O(X)$ . The circuit width can only grow a constant amount per stage, so its maximum width  $W \leq n + O(X)$ . So, fixing a circuit (which computes  $f$ ) minimal for  $C_{\min}$  and with its permutation count then optimized, the maximum extra cost in  $C'_{\min}$  due to permutations is  $p'(W)Y \leq p'(W)O(X)$ .

By the same argument as in problem 1, we have  $X \leq O(C_{\min}(f))$ , so this gives

$$\begin{aligned} C'_{\min}(f) &\leq C_{\min}(f) + p'(W)O(X) \\ &\leq C_{\min}(f) + p'(W)O(C_{\min}(f)) \\ &\leq C_{\min}(f)(1 + p'(W)O(1)) \\ &\leq C_{\min}(f)O(p'(W)) \end{aligned}$$

From our earlier bounds on  $W$  and  $X$ ,

$$W \leq n + O(C_{\min}(f))$$

If  $C_{\min}(f)$  is sublinear in  $n$ , this gives us  $W = O(n)$  and thus  $C'_{\min}(f) \leq C_{\min}(f)O(p'(n))$ , so we can pick some  $q = O(p')$ .

Otherwise we have  $W = O(C_{\min}(f))$  and

$$\begin{aligned} C'_{\min}(f) &\leq C_{\min}(f)O(p'(W)) \\ &= C_{\min}(f)O(p'(C_{\min}(f))) \\ &\leq q(C_{\min}(f)) \end{aligned}$$

for some polynomial  $q$  with  $\deg(q) \leq \deg(p') + 1 = O(\deg(p'))$ , as required. ■



### 4.7.3 adjoining a wire increases measured complexity

The worst way  $C_{M,\mathcal{P}}$  might behave “non-ideally” is “problem 3”, when adjoining a non-interacting wire can increase measured complexity of a generating circuit. Whether we can still rescue a lower bound on  $C_{\min}$  depends on how much of an increase we see.

If adjoining a wire always doubles measured complexity, it’s just what we’d expect from a random or naively constructed pattern basis (as explained in Section 2.13), which is also (not coincidentally) equivalent to treating the adjoined wire as a new input to a presumed “ $n$ -input gate” computed by the rest of the generating circuit. So there is no way to recover anything useful from this.

If it multiplies measured complexity by at most some smaller factor  $\alpha < 2$  (or in a more general model of computation, by a factor less than the number of possible states  $r$  of the adjoined wire), then extracting some bound might be possible; but I won’t pursue that here. (Whether this can be done is an open question, perhaps interesting — it’s conceivable that this would be compatible with proving some explicit function family’s circuit complexity to be superlinear in  $n$ . However, I have no reason to suspect that an attempt to come up with a useful pattern basis would find one with this kind of “intermediate level of usefulness”.)

On the other hand, if adjoining a wire only *adds* something to measured complexity, then we’re ok — even if each new wire adds some arbitrary polynomial  $p(n)$  in the circuit width  $n$ .

The analysis of this case is very similar to problem 2. But before diving into it, it’s worth looking at why this situation might plausibly arise while trying to find a useful pattern basis.

Recall from Section 2.13 that the way the adjoined-wire issue arises is that, in a naively constructed pattern basis, adjoining each wire tensors any function  $f$ ’s pattern matrix with itself (squaring the number of nonzero terms per row, thus doubling its measured complexity). The proposed fix (not yet proven possible) is a “change of basis” (at each inductive step, while constructing a pattern basis for the next higher value of  $n$ ), which would, we hope, reduce the number of nonzero terms back to what it was in the original (smaller) pattern matrix.

But we can imagine some strategy to reduce the number of nonzero terms (by “mixing” various combinations of naive-basis elements, to get single elements of the new basis) which succeeds partially but not completely. One way that could happen (probably not the only way) would be if the new terms were divided into  $P(n)$  classes (for some function  $P$  of the circuit width  $n$ ), with the mixing of the  $X^2$  nonzero “naive terms” resulting in at most  $X$  nonzero new terms per class, and therefore at most  $XP(n)$  nonzero new terms in all.

This is much better than the naive result of  $X^2$  nonzero terms (if there was no mixing, or if it didn't help), but much worse than the ideal result of only  $X$  nonzero new terms. But since  $X$  can be *doubly exponential* in  $n$  (i.e. have values up to  $2^{2^n}$ ), then for any  $P$  whose log is polynomial in  $n$  (say,  $P(n) = 2^{p(n)}$  for some polynomial  $p$ ), in terms of bounds this is closer to the ideal side than to the worst-case side — it results in only an *additive* measured complexity increase, namely an increase from  $(\log_2 X)$  to  $(\log_2 P(n) + \log_2 X) = (p(n) + \log_2 X)$ .

So in this sense, *any* significant partial progress towards a “useful” change of basis for the pattern matrix, as  $n$  goes to  $n + 1$  in a pattern basis being constructed by induction on  $n$ , might have a chance of reducing the growth of nonzero terms (from its natural growth due to “tensoring the pattern matrix with itself”) enough to allow the following analysis to apply.

So, what do we do if adjoining a wire to any  $g \in \mathcal{G}$  (of width  $n_g$ ) might increase its measured complexity  $C_{M,\mathcal{P}}(g)$  by an amount up to some polynomial  $p(n_g)$ ?<sup>52</sup> That is, what if we assume, for all  $g \in \mathcal{G}$ , when adjoining  $k \geq 0$  wires:

$$C_{M,\mathcal{P}}(\text{aw}_k(g)) \leq C_{M,\mathcal{P}}(g) + \sum_{i=n_g}^{n_g+k-1} p(i) \quad (8)$$

As before, we construct a modified  $C'_{\text{cost}}$  function, but this time we increase its value for almost every generating circuit  $g \in \mathcal{G}$  — whenever  $g = \text{aw}_k(g')$  and  $g' \in \mathcal{G}$ , we ensure

$$C'_{\text{cost}}(g) \geq C'_{\text{cost}}(g') + \sum_{i=n_{g'}}^{n_g-1} p(i) \quad (9)$$

by increasing  $C'_{\text{cost}}(g)$  however much is needed to accommodate all smaller  $g'$ .<sup>53</sup>

Assuming we're not worried about our bound being tight, the simplest way to do this is just to define, for all  $g \in \mathcal{G}$  (even for wire permutations),

$$C'_{\text{cost}}(g) = C_{\text{cost}}(g) + p'(n_g) \quad (10)$$

where for convenience we define

$$p'(n_g) = \sum_{i=0}^{n_g-1} p(i) \quad (11)$$

---

<sup>52</sup> We'll assume  $p(i) \geq 0$  for any  $i \geq 0$ .

<sup>53</sup> If we cared about a *tight* bound, we'd scale the added sum in (9) by some positive constant (dependent on  $g'$ ), which we'd optimize later.

which is easily seen to give  $C'_{\text{cost}}(g)$  the desired property (9).

We can also restate the assumed bound (8) using  $p'$ , to simplify its later use:

$$C_{M,\mathcal{P}}(g) \leq C_{M,\mathcal{P}}(g') + p'(n_g) - p'(n_{g'}) \quad (\text{when } g = \text{aw}_k(g') \text{ and } k \geq 0) \quad (12)$$

Of course the cost increase (10) makes problems 1 and 2 worse, but since it only adds a polynomial cost  $p'(n_g)$  to each generating circuit, we can still take care of those in the ways we did before. But the argument will be simplest if we handle all problems at once, so for simplicity we define a yet larger polynomial  $p''$  which can do that (and a corresponding  $C''_{\text{cost}}$ ), by also upperbounding the measured complexity of all free-in- $C_{\text{cost}}$  generating circuits (i.e. wire permutations and trivial wire operations, possibly with adjoined wires):

$$\forall_{g \in \mathcal{G}, C_{\text{cost}}(g)=0} : p''(n_g) \geq C_{M,\mathcal{P}}(g) \quad (13)$$

$$\forall_{n_g \geq 0} : p''(n_g) \geq p'(n_g) \geq 0 \quad (14)$$

$$C''_{\text{cost}}(g) = C_{\text{cost}}(g) + p''(n_g) \quad (15)$$

(We can't always define  $p''$  to satisfy (13), but we can if the preconditions for solving problems 1 and 2 are present.)

We again need a revised lemma to find  $\lambda$ :

**Lemma 9** *Given the assumptions and definitions above, there exists a  $\lambda$  with the same properties as in Lemma 5 (but for  $C''_{\text{cost}}$  rather than  $C_{\text{cost}}$ ).*

**Proof.** This time we have to minimize over an infinite set (in general) — namely, over *all*  $g \in \mathcal{G}$  with positive  $C_{M,\mathcal{P}}(g)$ . That makes the proof a bit more involved than for Lemmas 5 and 7 — besides showing  $C''_{\text{cost}}/C_{M,\mathcal{P}}$  is positive over a finite set of “base cases”, we have to show we made  $C''_{\text{cost}}$  grow fast enough (compared to  $C_{M,\mathcal{P}}$ ) to maintain a positive lower bound on that ratio as  $g$  gets arbitrarily wide.

Specifically, we set

$$\lambda = \inf_{g \in \mathcal{G}, C_{M,\mathcal{P}}(g) > 0} R(g) \quad (16)$$

where we define (for those  $g$ )

$$R(g) = C''_{\text{cost}}(g)/C_{M,\mathcal{P}}(g)$$

To prove  $\lambda > 0$  (from which the whole lemma follows as before), we must show:

- $C''_{\text{cost}}(g)$  is never 0 unless  $C_{M,\mathcal{P}}(g)$  is 0 — this follows from (15), (13), and  $p''(n_g) \geq 0$ . (Thus each  $R(g)$  used in the minimum is positive.)

- $R(g)$  has a positive lower bound for arbitrarily wide  $g$  — we'll prove this below, in two cases depending on whether  $C_{\text{cost}}(g) = 0$ .

When  $C_{\text{cost}}(g) = 0$ , (13) and (15) guarantee  $R(g) \geq 1$  (unless we skipped this  $g$  due to  $C_{M,\mathcal{P}}(g) = 0$ ).

Otherwise,  $g$  must be a possibly-widened primitive gate  $g'$ , so  $g = \text{aw}_k(g')$  for some  $g' \in \mathcal{G}_{\text{core}}$  and  $k \geq 0$ ; furthermore,  $C_{\text{cost}}(g) = C_{\text{cost}}(g') > 0$ . We'll divide such  $g$  into subsets  $B_{g'}$  (based on their primitive gate  $g'$ ), to get a collection of subsets, each internally indexed by  $k$ :

$$B_{g'} = \{\text{aw}_k(g') \mid k \geq 0\} \quad (\text{for } g' \in \mathcal{G}_{\text{core}}, C_{\text{cost}}(g') > 0)$$

Fixing  $g'$  and letting  $g$  stand for  $\text{aw}_k(g')$  whenever  $k$  is defined, we must show a positive lower bound on  $R(g)$  for  $g \in B_{g'}$  (whenever  $C_{M,\mathcal{P}}(g) > 0$ ); noting that the number of collections  $B_{g'}$  is finite will then complete the proof.

To do that, we expand  $R(g)$ 's numerator using (15),

$$C''_{\text{cost}}(g) = C_{\text{cost}}(g) + p''(n_g)$$

bound its denominator using (12) and (14),

$$\begin{aligned} 0 < C_{M,\mathcal{P}}(g) &\leq C_{M,\mathcal{P}}(g') + p'(n_g) - p'(n_{g'}) \\ &\leq C_{M,\mathcal{P}}(g') + p'(n_g) \\ &\leq C_{M,\mathcal{P}}(g') + p''(n_g) \end{aligned}$$

and combine these (recalling  $C_{\text{cost}}(g) = C_{\text{cost}}(g')$ ) to get

$$R(g) \geq \frac{C_{\text{cost}}(g') + p''(n_g)}{C_{M,\mathcal{P}}(g') + p''(n_g)}$$

(whose only dependence on  $k$  comes implicitly from  $n_g = k + n_{g'}$ ).

If  $C_{\text{cost}}(g') \geq C_{M,\mathcal{P}}(g')$  then  $R(g) \geq 1$  (even if  $C_{M,\mathcal{P}}(g') = 0$ , which we can't rule out here); otherwise  $0 < C_{\text{cost}}(g') < C_{M,\mathcal{P}}(g')$  and (recalling  $p''(n_g) \geq 0$ ) we easily derive

$$R(g) \geq C_{\text{cost}}(g')/C_{M,\mathcal{P}}(g') > 0$$

which completes the proof. ■

The  $\lambda$  from Lemma 9 will scale  $C_{M,\mathcal{P}}$  so Theorem 4 can prove  $C_{M,\mathcal{P},\lambda}$  lowerbounds  $C''_{\min}$  (which we define in the usual way based on  $C''_{\text{cost}}$ ), just as in problem 2. What

remains to establish (by proving the following Proposition) is almost identical to what it was for that problem — though as with problem 1, we need to add  $n$  and  $m$  to  $C_{\min}(f)$  for our conclusion to remain correct for those  $f$  which require mostly trivial wire operations to compute.

**Proposition 10** *Given the assumptions and definitions above (particularly the definition of  $C''_{\text{cost}}(g)$  in (15), but also (11)–(14) about the fixed polynomials  $p$ ,  $p'$ , and  $p''$ ), there is a polynomial  $q$  such that, for all  $f \in \mathcal{F}$  (with  $f: \mathbb{F}_2^n \rightarrow \mathbb{F}_2^m$  for any  $n$  and  $m$ ),  $C''_{\min}(f) \leq q(C_{\min}(f) + n + m)$ . Furthermore,  $q$  is independent of  $n$  and  $m$ , and  $\deg(q) = O(\deg(p''))$ . (And  $\deg(p'')$  is no more than a constant plus the degree of any polynomial bound regarding  $C_{M,p}$ , over various subsets of  $\mathcal{G}$ , which we had to accomodate when defining  $p''$  above.)*

**Proof.** This is also very similar to problem 2, except this time *every* generating circuit has increased cost. (Its new statements about polynomial degrees are obvious.)

As before, choose a minimal circuit for  $f$  according to  $C_{\min}$ , and say it has  $X$  non-free generating circuits (according to  $C_{\text{cost}}$ , which means they're all perhaps-widened primitive gates); note that it's possible  $X = 0$ . Then holding the core circuits inside those  $X$  generating circuits fixed (including their ordering and input signals), revise the circuit as a whole to minimize its number  $Y$  of free (according to  $C_{\text{cost}}$ ) generating circuits (which will of course not change its cost). Combining arguments from problems 2 and 1 (about wire permutations and trivial wire operations respectively, which are the only free generating circuits allowed by  $C_{\text{cost}}$ ), we know  $Y \leq O(X) + n + m$ . (We fix the resulting circuit for the rest of the proof.)

Also repeating arguments from other problems, we can upperbound  $X \leq O(C_{\min}(f))$ , and the maximum width of any circuit stage (of which there are  $X + Y$  in all):

$$\begin{aligned} W &\leq n + O(X + Y) \\ &\leq n + O(C_{\min}(f)) + O(X) + O(n + m) \\ &\leq O(C_{\min}(f)) + O(n + m) \\ &\leq O(C_{\min}(f) + n + m) \end{aligned}$$

This lets us (loosely) bound the extra cost seen by  $C''_{\min}$  over  $C_{\min}$ , since this is at most  $p''(W)$  for any circuit stage (note that this covers the extra cost due to all 3 problems we're considering):

$$\begin{aligned} C''_{\min}(f) &\leq C_{\min}(f) + p''(W)(X + Y) \\ &\leq C_{\min}(f) + p''(O(C_{\min}(f) + n + m))(O(C_{\min}(f)) + n + m) \\ &\leq q(C_{\min}(f) + n + m) \end{aligned}$$

for some polynomial  $q$  with  $\deg(q) \leq \deg(p'') + 1$ . ■

## 5 Motivation

This section presents heuristic arguments which, though they’re not proofs and some of them refer to vague concepts, seem to the author to be evidence that if circuit complexity can be understood in some way similar to the conventional intuitions about related phenomena, something like the approach and ideas summarized in Section 2 ought to be a good way to try to do it. (These are also a condensed version of some of the lines of reasoning that led to that proposal.)

These arguments are not *part* of that proposal, but are presented in the hope of making its usefulness (in guiding further work) seem more likely, at least to someone who shares the author’s belief that *some* way of understanding this phenomenon of circuit complexity must exist.

### 5.1 the “pattern intuition”

A conventional intuition about computational complexity imagines that certain functions  $f$  contain “inherent computational work” which any process that computes them must “do in some form” (but which a single gate could only do a small amount of), and furthermore that this requirement can in some cases be perceived in the form of “patterns in the truth table of  $f$ ”. (Indeed, a typical “naive false proof” of  $P \neq NP$  consists essentially of asserting this intuition as fact. [Tre08])

The hard part, of course, includes formalizing “computational work” and “do in some form” and “pattern”, including understanding in what sense “computational work” can be localized to specific combinations of gates in a circuit, and proving the resulting formal concepts have the necessary properties. Evidently those properties would need to include something like:

1. the work that must be done to compute a function depends on the patterns in its truth table;
2. independent work must be done independently, so a certain number of gates can only do a certain amount of it — though there doesn’t need to be a direct correspondence between single gates and single pieces of work; that is, some kind of “superposition” of multiple pieces of work done by a single primitive computation is acceptable, and some work might be done by correlation of activities in multiple places, as long as there remains some quantitative limitation on how much work can be done by a given amount of computation;
3. when composing functions like  $f = g \circ h$ , patterns in  $g$  and  $h$  combine to make patterns in  $f$ ;

4. that is the *only* way patterns can exist in  $g \circ h$ .

(Simple attempts to formalize those concepts encounter serious difficulties, most of which needn't be elaborated here; others will be mentioned below when they come up.)

The latter rule seems essential to proving any complexity bound, since otherwise “patterns could come out of nowhere” (that is, you couldn't prove they couldn't), but it's severely tested by the existence of simple pseudorandom functions which can “reversibly encrypt and decrypt” their input bit vectors. Looking at a circuit composed of successive pieces (subcircuits)  $s_i$ , which (for example if  $i \in \{1, 2, 3\}$ ) computes the function  $f = s_3 \circ s_2 \circ s_1$ , we can view it “backwards in time” as an operation of each piece, on the truth table of the rest of the circuit, to produce the truth table of the portion of the circuit starting with that piece. So, for example,  $s_2$  modifies the truth table of  $s_3$  to produce that of  $s_3 \circ s_2$ , and similarly  $s_1$  maps  $s_3 \circ s_2$  to  $s_3 \circ s_2 \circ s_1$ . Then if (forwards in time)  $s_1$  encrypts its input,  $s_2$  decrypts that, and  $s_3$  computes something interesting, in the backwards view we see  $s_3$  with obvious patterns in its truth table,  $s_3 \circ s_2$  with no obvious patterns, but  $s_3 \circ s_2 \circ s_1 = s_3$  and has the original patterns revealed again.

If the pattern intuition can be saved, it seems necessary to hypothesize that  $s_3 \circ s_2$  contains “*hidden patterns*” rather than no patterns, which can be revealed again by composing it with  $s_1$  (especially if  $s_3$  has much higher complexity than either  $s_1$  or  $s_2$ , so it seems impossible that  $s_1$  “recreates the patterns from nothing”).

## 5.2 can linear algebra help?

Is there any precedent in mathematics for some kind of formal pattern with properties 3 and 4 in the above list? Yes — linear maps between vector spaces, which (given a choice of basis in each space) can be concretely represented as matrices and composed using matrix multiplication, with a nonzero value of a matrix entry viewed as “presence of a pattern” (to a degree corresponding to its absolute value or magnitude).

And it turns out that composition of general boolean functions *can* be represented linearly this way, in which a function  $f$  mapping  $\mathbb{F}_2^n \rightarrow \mathbb{F}_2^m$  can be represented as a  $2^m$  by  $2^n$  matrix of real-valued ***fourier coefficients*** (most conveniently computed if we represent a bit  $b$  in  $\mathbb{F}_2$  by the power  $(-1)^b$ , i.e. use 1 and -1 in place of 0 and 1; each coefficient is just the expected product of some XOR of inputs and some XOR of outputs (also represented by  $(-1)^b$ ) over input vectors chosen uniformly at

random in  $\mathbb{F}_2^n$ ).<sup>54</sup> Then if  $f = g \circ h$  and  $\text{FM}()$  turns a boolean function into the corresponding  $2^m$  by  $2^n$  *fourier matrix*, we have simply  $\text{FM}(f) = \text{FM}(g) \text{FM}(h)$ . (This “*fourier matrix model of boolean computation*” was discussed more extensively in Section 2.16.3.)

The “patterns” in this view then consist of correlations between XORs of subsets of outputs and XORs of subsets of inputs, and this linear representation makes it clear that such patterns can never be “created out of nothing” — unlike correlations between single inputs and single “intermediate wires”, which (as we follow some sequence of causally related wires forward in a circuit) can go completely to zero after XORing some input or intermediate wire with uncorrelated data, and nonetheless come back into existence by later XORing that wire with the same uncorrelated data (restoring its original *signal*, i.e. whatever function of the original inputs specifies its value).

Unfortunately that doesn’t solve our problem — even though encryption or decryption (and all other computation) can only move correlations around linearly rather than creating them, it can:

- do so in a way that doesn’t obviously conserve some kind of “total magnitude” (since a phenomenon of “interference” can occur, which can be constructive, at least when using the standard euclidean norm to judge magnitudes, which is natural due to the sum of squares of any signal’s fourier coefficients being 1);
- efficiently spread them into so many XORs of different subsets of intermediate wires that they are apparently undetectable as being significant — thus the distribution of fourier coefficients of a pseudorandom function “looks the same” as that of a random function. (And in hindsight, the natural proofs barrier implies we should have expected that — computing all such correlations from the truth table is too easy to be able to say much about complexity.)

This “spreading thin” of potentially significant independent signals, together with the exponential number of potential “channels” for information (namely, each distinct XOR of intermediate wires), also makes it unclear that there is any limit to how much “work” one gate can do, when it acts on a set of input wires containing (in their correlations by XOR with various combinations of intermediate wires anywhere in the circuit) contributions from many signals.

(Combinations of fourier coefficients can in principle give complete information about the function, but the same can be said for single bits in the truth table, so this

---

<sup>54</sup> By *XOR* or “exclusive or” we mean addition in  $\mathbb{F}_2$ ; this is represented by multiplication when we represent bits by  $(-1)^b$ .



is not very relevant. Note that in terms of  $n$ , there are the same number of truth table rows and fourier matrix columns, namely  $2^n$ ; the apparent difference between  $m$  truth table columns vs  $2^m$  fourier matrix rows goes away when we account for each XOR of outputs, not just each output alone, which is necessary if we want function composition to be bilinear.)

So if the pattern intuition is to be saved, evidently the kinds of patterns need to be more complex than just correlations between XORs of wires (including inputs and outputs) — even though this will make the function overdetermined by the complete set of pattern values.

### 5.3 higher-degree patterns

So we look for more kinds of patterns. One thing the fourier coefficients and single truth table bits have in common is that they are degree-1 polynomials in the truth table entries (again represented as  $(-1)^b \in \mathbb{R}$ ). (In fact, either set taken alone is exactly a  $2^n$ -element *basis* of all (homogeneous) degree-1 polynomials in those entries, for a single-output truth table. Note that powers higher than one are not useful in formal real polynomials whose variables correspond to truth table entries using that representation, since each entry  $v \in \{1, -1\}$  obeys the relation  $v^2 = 1$ .)

What about higher-degree polynomials in the truth table entries? (We'll focus on the  $m = 1$  (single-output) case here, for simplicity and since it seems representative of the phenomena we need to explain.) Several observations make higher degrees potentially interesting:

- If we try to quantify the amount of “interference” that affects the linear propagation of degree-1 polynomials through reversible gates such as S-boxes (a typical building block of reversible encryption), we find degree-2 polynomials that measure it. Similar things can be said about higher degrees.
- The degree- $k$  polynomials, as a set, also propagate linearly through reversible functions. There are more of them, so maybe they can say more about a function than the degree-1 polynomials can. (Their behavior is not so simple through irreversible functions — it's still linear, but only when other degrees are included, since terms can turn into lower-degree terms (of the same parity of degree) due to multiple truth table entries in the input space  $\mathbb{F}_2^n$  mapping to one entry in the output space  $\mathbb{F}_2^m$ , combined with the relation  $v^2 = 1$  obeyed by every entry. But it seems clear that even restricting our attention to reversible functions, we encounter most of the same phenomena that need to be

understood; and any computation can be done reversibly (except for erasing or forgetting the input, and copying the final output) with at most a small increase in circuit complexity [Ben73]. So we’re free to focus temporarily on reversible functions, if it brings some advantage of simplicity.)

- Certain degree- $k$  polynomials (for  $k = 2^d$  and  $2^{d-1}$ ) have a close relationship with measuring properties of simple functions consisting of  $d$ -input AND gates (possibly in parallel with  $\mathbb{F}_2$ -affine functions on other wires), or arbitrary XORs of such functions. (Details unpublished, since this is not important for now, and simple enough that it’s probably well-known.)
- The  $2^d$ th power of the  $d$ th Gowers uniformity norm<sup>55</sup> is related to the polynomials just mentioned; that norm also has interesting properties [TZ11].
- The “influence” of a set of  $d$  inputs [KKL88] is also a degree- $2^d$  polynomial.
- Products of different degree-1 polynomials also have interesting properties. Using these and Taylor series, we can come up with polynomials that can be interpreted as analyzing how well any function  $f$  matches “an approximate version of a given function of a specified subset of its inputs” (or even, “of a specified set of XORs of subsets of its inputs”), which is the kind of property that intuitively seems to imply the need to do specific “computational work”.
- In fact, even just looking at products of powers of the  $2^n$  (degree-1) Fourier coefficients (which are directly related to the properties just mentioned about approximating simple functions), we find interesting phenomena (at least when  $K$  is  $\mathbb{R}$  or  $\mathbb{C}$ , which we’ll assume in this paragraph). Unlike the Fourier coefficients taken alone, products of their powers are sufficient to span the entire property space (proving this is a simple but instructive exercise). That means the property space has a basis consisting only of such functions. (Whether such a basis might be a “useful pattern basis” is unknown, but I doubt it.) But I know of no simple description of any example of a basis like that — these products have complicated relations of linear dependence in the property space.

The problem with investigating more and more interesting kinds of properties is that we find no natural “end” to the set of properties we seem to need to include, if we want to track the effect on simple properties of being evaluated on the composition of more and more S-boxes. For example, even if their degree stays fixed, their

---

<sup>55</sup> For more info and references, see Section 2.16.2.

“complexity” (in terms of how their polynomial coefficients relate to the structure of the truth table entry indices) increases, due to the “mixing” by the S-boxes.

To get ahead of myself, this problem was eventually resolved by just tracking *all* properties, i.e. by treating functions as linear maps between complete property spaces. This didn’t so much answer the question as reformulate it, since the important issue then became finding the right basis for those spaces. But it did potentially solve the important problem of how to prove a complexity lower bound.

But for a long time it seemed like it was instead necessary to find some subset (presumably, in hindsight, one which didn’t span the property space) of especially important properties which ought to be tracked, and which, combined, would be sufficient in some not-yet-understood way to prove a complexity bound. So in case any readers might still be wondering if that way wouldn’t end up being better, I want to continue in this Motivation section (as I did in reality) with a line of reasoning which doesn’t yet assume we need the whole property space (and doesn’t yet explicitly understand that properties should be thought of as part of that space at all), but does examine the question of just how many properties we want to track, and include as part of a complexity formula.

## 5.4 pseudorandom vs random

At this point we can revisit the kind of model we were looking for, a bit more formally:

- we would like to find a linear model in which matrix entries act like patterns in functions (that is, in their truth tables), and must be nonzero to map “patterns in the function applied to the output bits” (i.e. in a truth table visible to the outputs) to “patterns in the function applied to the input bits” (i.e. in the truth table thereby induced on the inputs);
- we don’t yet know whether we can do that at all, since the properties we want to track might not be “linearly closed” from the point of view of predicting their values in  $g \circ h$  from those in  $g$  and  $h$  (even for small specially chosen  $h$ , which would be sufficient) (and for some reason that worry didn’t immediately lead to the idea of “just linearly closing them” by forming and using the whole property space); so in place of that, we hope to find “some sort of inequality” relating the values of the properties we want to track before and after composition;
- but in any case, we expect to need more (and more complex) patterns than just fourier coefficients, which means the matrix size (or just the number of patterns, if there is no matrix) will be higher than  $2^m$  by  $2^n$ , and the function will be overdetermined by its pattern values.

- (It also seems sufficient to restrict attention to looking for patterns in single-output functions ( $m = 1$ ), except when we’re focussing on reversible functions (which requires  $m = n$ ), for the same reason that in  $g \circ h$  we can assume  $h$  is small and specially chosen (with  $h$ ’s  $m$  and  $n$  general, but near in value, since  $h$  can consist of one simple gate and any number of parallel wires) — namely, that the phenomena needing explanation come up even when  $m = 1$ , and all circuits can be built by composing them in stages (from end to beginning) from such  $h$ . But some thought experiments require considering reversible functions, and a general theory would seem incomplete if it couldn’t handle any  $h$  in  $g \circ h$  and compose functions based on any  $(n, m)$ , so we don’t want to ignore general numbers of outputs entirely.)

In those terms we can restate how we hope pseudorandom functions will be handled (whether they’re reversible, or have one output, or are fully general):

- pseudorandom functions do have patterns, but only complex patterns — not readily-apparent simple patterns, such as fourier coefficients with large absolute values, or even simple low-degree highly-structured polynomials in the truth table entries (with large absolute values).

What about random functions — to be most definite, those with near-maximal Kolmogorov complexity (which includes most functions)?

Even those could be used (very inefficiently) for reversible encryption, so just like pseudorandom functions, they must have (and induce, in their compositions with functions of interest) hidden patterns rather than no patterns. Not only that — since we want to prove “functions satisfying more patterns must be more complex”, random functions must satisfy *more* patterns than any other kind of function.

But this leads to an apparent paradox (also mentioned in the Summary): assume  $f$  has near-maximal Kolmogorov complexity. Suppose  $f$  satisfies pattern  $p$  (now treating  $p$  as a predicate — if patterns are matrix entries, we just mean by this that the entry has magnitude higher than some threshold). Suppose further that  $p$  is a ***discriminating*** pattern, i.e. only a small fraction of all functions (say  $2^{-k}$ ) satisfy  $p$ . Then the fact that  $f$  matches  $p$  would allow us to describe  $f$  using almost  $k$  fewer bits than in its truth table, contradicting our assumption about  $f$  — unless  $p$  itself requires almost  $k$  bits to describe.

In other words, for a “random”  $f$ , any (predicate) pattern it satisfies must itself require almost as much information to describe, as its satisfaction implies about  $f$ . (And since for purposes of Kolmogorov complexity, a pattern could be described by

its position in our list of patterns, that also means we need a high *number* of patterns if we want *any* of them to be discriminating and apply to random functions.)

But do the patterns *need* to be highly discriminating? (If they do, surely we need a very lot of them, comparable at least to something like  $(2^{2^n})^\epsilon$ , both for the reason just given, and just to make sure there are still enough of them matching each random function so that its complexity can be measured high enough — since the basic idea of getting patterns in  $f = g \circ h$  from combinations of patterns in  $g$  and in  $h$  does imply some sort of “multiplication”, and therefore that a function of complexity  $X$  should have a number of patterns more like  $c^X$  than like  $cX$ . (I also had other reasons, not yet written up, for thinking of pattern combinations as “inherently multiplicative” — for example, given pattern  $p1$  in circuit  $g1$  and  $p2$  in  $g2$ , that adjoining those circuits in parallel (in a non-interacting way) should lead (for each pair of  $p1$  and  $p2$  in those circuits) to a “joint pattern” or “combined pattern” in the adjoined circuit, which we might think of as  $p1 \otimes p2$ . As it turned out later, that is some even fancier operation than the tensor product in their property map spaces — so we might call it a “star product”  $p1 \circledast p2$  if it ever needs naming, reserving  $\otimes$  for the standard tensor product — but it’s still something multiplicative rather than additive.))

That is where the “random probing” argument from the Summary comes in, which suggests that the natural proofs barrier implies that almost all of the pattern values for *any* function need to be either zero or “insignificantly small” — so they are not a significant source of measured complexity of a function, so that polynomial-time random probing can’t estimate its complexity — which means only a relatively few patterns must be the source of all its measured complexity. And yet (1) those “few patterns” must still number around  $c^X$ , and (2) this means most pattern values are zero or insignificant, and therefore most patterns are highly discriminating, and therefore (reiterating a point above) many patterns are highly complex.

(For completeness, I should dispense with an alternate possible interpretation of that “random probing” argument — maybe many single patterns take a superpolynomial time in the truth table size to compute, so that argument doesn’t actually show there must be a superpolynomial ratio of nonmatching to matching patterns. I can’t strictly rule that out, but I consider it unlikely — not only do we know of many simple patterns which are both fast to compute and evidently significant in their effect on compositions, but the very fact of a pattern being significant in effect on compositions seems to indicate it should have “intrinsic complexity” (a vague term, but more like computational than Kolmogorov complexity) more like  $2^n$  (like a truth table) or lower, than like  $2^{2^n}$  (like an arbitrary predicate-like property of truth tables), for intuitive reasons I can’t quite pin down — probably some belief about

the *nature* of a pattern like that, e.g. that it measures (both the presence and the “direction” or “sign” of) some special kind of “asymmetry” in the truth table, and those kinds of asymmetry can’t be too complicated and still make sense.)

Putting all this together, there seems to be no way to avoid needing a number of patterns roughly comparable to the number of truth tables, in the sense of being something like  $(2^{2^n})^\epsilon$ , preferably with  $\epsilon$  being reasonably large (like at least  $1/2$ ). (And as we know, I later ended up with  $\epsilon = 1$ .)

## 5.5 but with so many patterns, what can they be like?

But then, what kind of pattern, of the level of complexity that implies, can possibly “make sense” as part of a complexity formula, so that the fact of its matching some function  $f$  “means something” — either by making it likely that compositions like  $f \circ h$  match other (perhaps similar or generally simpler) patterns,<sup>56</sup> or by giving *direct* evidence that  $f$  is hard to compute? (By “direct evidence”, I meant evidence that *doesn’t assume* that matching more patterns implies higher complexity — since the idea here was to *use* that “direct evidence” (from lots of independent patterns, combined) to *help prove* that matching more patterns implies higher complexity. Possible examples of this kind of evidence might have been: proof that “functions that match  $p$  are, on average, harder to compute than functions that don’t match  $p$ ”; or evidence that  $f$  matching  $p$  implies something about the structure of any computational process resulting in  $f$ , which is both nontrivial, and independent of similar implications due to matching other patterns.)

The short answer is — “I don’t know.” (I’ve had lots of ideas, a few of which might turn out to be relevant to finding a useful pattern basis; but most of those are either mentioned in this paper, or too vague or incomplete to write down.)

But I do know that it’s hard to find any list of patterns that seems complete enough, unless it provably spans the property space — since otherwise, chasing the patterns through composed S-boxes invariably leads to patterns not linearly spanned by patterns in the list, with no obvious way to prove any inequalities relating them to the ones in the list. And in hindsight, once all patterns are viewed as part of the property space, and all boolean functions are viewed as linear maps between such

---

<sup>56</sup> More to the point, in light of our earlier “rule 4” about the only source of patterns in a composition like  $f \circ h$  being patterns in the components, would be patterns in  $f$  and in  $h$  somehow being necessary to “allow” patterns to exist in  $f \circ h$  (though exactly what that could mean was very unclear). One advantage of focusing on reversible  $h$  was to remove this distinction (between the two possible directions of implication between patterns in  $f$  vs. patterns in  $f \circ h$ ). In the property space model, this issue is resolved by function composition (or more precisely, the induced property map composition) being bilinear, so this reason for focusing on reversible  $h$  is no longer relevant.

spaces, and noting that reversible boolean functions can map simple patterns to a huge variety of complex properties — and then their inverses can precisely undo that, reconstructing the original patterns perfectly — it seems both unwise and unjustified to leave *any* properties out of the linear span of “all patterns in our list” (when we want to list the ones whose linear evolution through composed S-boxes we need to understand, so they can end up being included in our complexity formula).

Another way to say essentially the same thing: you’ll never be able to prove that some arbitrary pattern (or a sum of many patterns) can’t show up out of the sequence  $decrypt \circ do' \circ encrypt$  for a simple circuit  $do'$ , even if trying to get it directly out of a straightforward circuit  $do$  would be hard (i.e. require a complex circuit), unless you can take into account all the patterns that might appear in the intermediate stages, even though they are “hidden patterns” whose form we don’t yet specifically understand — but those might be arbitrarily complex, if we want to do this for complex *encrypt / decrypt* pairs. So although it seems unlikely that there can be a version of “homomorphic encryption” which makes certain complex functions (unrelated to the encryption scheme) *easier* to compute rather than harder, to prove it *can’t* happen you’d better be able to account for *all* the patterns involved in encrypted or pseudorandom functions. (And there doesn’t seem to be any way to do that without in some sense understanding a complete linear basis of all properties — certainly you can’t hope to understand them individually, but that way you might be able to understand something about the whole set.)

So given that, *and* the need for almost as many properties, and almost as complex properties, as there would be in a complete basis of the property space, *and* the fact that once you have the *whole* property space you can potentially use almost any submultiplicative matrix measure to prove the basic theorem you’d need to lowerbound complexity (rather than trying to find “some sort of inequality we don’t yet understand”) — there doesn’t seem to be any reason *not* to use it.

## 5.6 the property space has room for $2^{(\text{number of gates})}$ patterns

In reality there were several other threads to my motivations, besides the ones summarized above. What comes to mind (without delving into my notes) includes: considerations of “multiplicativeness of patterns” (already mentioned) and other thought experiments about how patterns might be localized (or not) in a circuit and might combine when circuits combine; information theory and cryptography; various analogies to physics and quantum computing; lots of exploration of potential specific patterns and mathematical topics that might relate to them (like root systems, representations of the symmetric group, and finite fields); and probably others.

Of the topics I listed which seem clearly relevant in hindsight, I think this relevance is apparent enough in the current proposal to not need further discussion here.

But one thing I should elaborate on is the vision I ended up with (which helped lead to, and remains compatible with, the present proposal) of sums of dual patterns moving (backwards in time) through circuit stages consisting of S-boxes, with each stage turning each dual pattern into a small sum of dual patterns (due to the small number of nonzero pattern values on any pattern matrix row for that stage). (Recall that a “dual pattern” is just a formal linear combination of truth tables; informally I sometimes just call these “patterns”, especially in the context of reversible functions, so I may do that in some places below.)

One reason is that this is a good mental image for thinking about some of the important analogies already mentioned, and especially about the applicability of linear algebra (and related phenomena like “interference”) to understanding this evolution of patterns. (It seems likely that further exploration of those analogies, for example the analogy to wavefunctions and linear operators in quantum mechanics, could help provide ideas for the kinds of patterns needed in a useful pattern basis. Note that this is *not* already being done (directly or consciously, anyway) by the theory of quantum computation (or, as far as I know, by any kind of conventional quantum physics), since that is “one level of exponent lower” (even when it considers infinite-dimensional spaces of functions), since it considers wavefunctions over physical configurations, rather than wavefunctions over *arbitrary functions* of physical configurations (as we do here, since those arbitrary functions are analogous to truth tables). For the same reason, if this framework is generalized to apply also to quantum computation (which I have not attempted here), the wavefunctions involved in that computation will probably be of a distinct kind and level from the ones involved in this analogy, though they might have some relationship — whether and how they relate will be interesting to find out.)

But the main reason that vision (of an S-box transforming one (dual) pattern into a small sum of them) is relevant to this Motivation section is its idea of limiting that sum to a *small* sum — both that it might be possible (even when it involves complex patterns — or, perhaps, *because* it does, given the challenge of maintaining the smallness as each adjoined parallel wire tensors the gate’s pattern matrix with itself); and, if it *is* possible, that that would be sufficient to prove a complexity lower bound, since it would prove that any sum of patterns can only increase *exponentially* in number of terms, as a circuit gets more complex. (It took awhile to get used to considering a “merely exponential growth rate” as desirably small, but you can do that when you’re working in a space of dimension  $2^{2^n}$ .)

This is crucial, since it removes the need for a pattern to be *intrinsically* complexity-



proving — it’s sufficient instead to just prove that you can only get to it (from the simple patterns you start with) in a limited number of ways — i.e. only by paths through the few nonzero pattern values in the simple pattern matrices, one per gate, which are multiplied to make the whole circuit’s pattern matrix. And that means (since the total “number of ways” you can get to *any* pattern is then limited too, if you can find a useful pattern basis in the sense described by this vision) that each gate can only multiply your pattern count by some constant, which is enough to get you the necessary result — to lowerbound complexity by (roughly) the log of the pattern count.<sup>57</sup>

(And putting all this into the framework of a submultiplicative matrix measure means you might have a way to prove that result, *even if you can’t keep the number of nonzero entries small*, provided you can keep the number of non-“small” entries small (on each row), and find a more sophisticated matrix measure which is not bothered by the “small” entries.)

## 6 Discussion

### 6.1 But does it “naturalize”?

As we discussed briefly in the Summary (in Section 2.3), the “*natural proofs barrier*” of Razborov & Rudich [RR97] refers to the fact that, under assumptions about pseudorandomness which are generally believed, any property  $P$  (of the truth table of an arbitrary function  $f_n$  from  $n$  bits to 1 bit) used to lowerbound asymptotic circuit complexity of some function family  $\langle f_n \rangle$ , by provably containing each  $f_n$  but no “simple” functions, can’t be both “*large*” (apply to a non-negligible fraction of functions of each size  $n$ ) and “*constructive*” (be computable in polynomial time<sup>58</sup> in the size of  $f$ ’s truth table). (For a more precise statement, see [RR97]. The terms “large”, “constructive”, and “natural” have technical meanings within that paper

---

<sup>57</sup> As discussed in Section 4.7.3, for some purposes it would suffice to prove only that each new gate multiplies the circuit’s pattern count by no more than  $2^{p(n)}$ , where  $n$  is the circuit *width* (wire count) at that gate’s stage, and  $p$  is any fixed polynomial; this still lets you lowerbound some fixed polynomial of circuit complexity by (roughly) the log of the pattern count, which could still let you prove a superpolynomial lower bound.

<sup>58</sup> In fact, [RR97] comments that “constructive” can be generalized from “computable in polynomial time” to “computable with quasi-polynomial-sized circuits” (applied to  $f$ ’s truth table), but this doesn’t significantly affect the present discussion, since the properties we’ll consider below (for at least some plausible parameters of our proposed framework) only become obviously computable by circuits if we allow those to have exponential size in  $2^n$  (i.e.  $(2^{2^n})^\epsilon$  for some  $\epsilon > 0$ ), since they have a “query complexity” of that size in the pattern magnitudes of  $f$ .

(which we tried to summarize accurately above); we’ll use quotes here when using them in that technical sense. The term “useful” also has a technical meaning in [RR97], but we don’t need to refer to that explicitly here, so we only use that term in *this* paper’s technical (albeit partially informal) sense (about a “useful” pattern basis).)

The first test of whether this barrier might rule out a potential proof strategy of the type we’re proposing here — defining a “measured complexity formula”  $C_{M,\mathcal{P},\lambda}$  which lowerbounds the circuit complexity of any function  $f_n$  (with input size  $n$ ) — is simply whether the property  $P$  defined as “ $C_{M,\mathcal{P},\lambda}(f_n) \geq X_n$ ” (for an  $X_n$  which expresses the lower bound we want to prove, as a function of  $n$ ) is both “large” and “constructive”.

From our earlier discussion of the kind of “useful pattern basis” we hope can be found, clearly that property would be “large” — in fact, it would apply not only to a “non-negligible fraction of functions” (for each  $n$ ), but to *most* functions.

On the other hand, deciding  $P$  in a brute force way would involve summing  $2^{2^n}$  terms, so it’s reasonable to hope  $P$  would not be “constructive”. (As discussed in Sections 2.3 and 2.16.1, even *guessing*  $P(f_n)$  by sampling some fraction of terms, when promised  $f_n$  is either random or “pseudorandom”, would take too long for this, since so many terms would be 0 that we’d rarely examine any nonzero terms.)

However, the natural proofs barrier has a subtlety — a typical proof that each  $f_n$  has property  $P$  is likely to identify some more specific property  $P'$  (which is, more or less, “whatever is special about  $f_n$  which lets us prove it has property  $P$ ”), and then prove “each  $f_n$  has  $P'$ , and  $P'$  implies  $P$ ”. Then it’s not enough for  $P$  itself to avoid being both “large” and “constructive” —  $P'$  must meet that test as well.<sup>59</sup> And if there’s any ambiguity about how to define  $P'$ , *all* ways of doing it must pass this test. Otherwise, whichever definition of  $P'$  doesn’t pass it allows the barrier to show that we’ll never succeed in making the whole proof work (unless our assumptions about pseudorandomness are wrong, which (like others) I consider unlikely). (When a  $P'$  that fails this test can be found, the proof attempt that explicitly involved  $P$  is said to “**naturalize**”. Thus, we’re hoping to argue here that the present proposal doesn’t inevitably “naturalize”.)

So, let’s examine whether the present proposal meets this more difficult test. (Of course we won’t be able to *prove* it does (short of successfully using it to prove a significant lower bound); but at least we can try to rule out some obvious candidates for a “naturalizing property”  $P'$ .)

---

<sup>59</sup> This subtlety was fully discussed in the original paper [RR97], and also in some truly marvellous “popular expositions” whose URLs the author’s memory appears to be too unreliable to retain.

Suppose we found a pattern basis  $\mathcal{P}$  and associated complexity measure  $C_{M,\mathcal{P},\lambda}$  which were “useful” in the sense discussed in the Summary, and then proved that some specific function family  $\langle f_n \rangle \in \text{NP}$  had superpolynomial measured complexity according to  $C_{M,\mathcal{P},\lambda}$  (and therefore, superpolynomial actual circuit complexity, meaning it was not in  $\text{P/poly}$ ).

We saw above that the boolean property<sup>60</sup>

$$P(f) = (C_{M,\mathcal{P},\lambda}(f) \geq X_{n_f})$$

would be “large” but presumably not “constructive”. ( $\langle X_n \rangle$  expresses our superpolynomial bound as a function of  $n$ ;  $n_f$  denotes the input size of any function  $f$ .)

To look for the kind of smaller property  $P' \subset P$  that we ought to worry about, recall the basic idea in the definition of  $C_{M,\mathcal{P},\lambda}(f)$  — roughly, it’s the log of the sum of  $2^{2^n}$  terms, each one the magnitude of a “pattern match value” present as a matrix entry in (some row of)  $f$ ’s pattern matrix  $\text{PM}_{\mathcal{P}}(f)$ .<sup>61</sup> For any specific  $f_n \in \langle f_n \rangle$ , a relatively few of those “pattern magnitudes” have high values (call them the set  $Y_n$ ), with the remainder having 0 or insignificant value. And the kind of proof of high complexity we anticipate is simply to identify enough of those  $Y_n$  (given our specific knowledge of  $f_n$ ), and prove they have a reasonably high sum when taken alone.

Given this, the obvious choice for a more specific property  $P'$ , which applies to a specific function family  $\langle f_n \rangle$  and implies  $P$ , is just that the part of  $C_{M,\mathcal{P},\lambda}(f_n)$  which comes from the pattern magnitudes in  $Y_n$  (a subset of the pattern basis set  $\mathcal{P}_{P_n}$ ) is high:

$$P'(f) = (C_{M,\mathcal{P}(Y_n),\lambda}(f) \geq X_{n_f})$$

(where the notation  $C_{M,\mathcal{P}(Y_n),\lambda}(f)$  is made up just for this section, with the meaning of ignoring all patterns not in  $Y_n$  when evaluating  $C_{M,\mathcal{P},\lambda}(f)$  on any  $f$  with  $n$  inputs).

So, is *this*  $P'$  both “large” and “constructive”?

Whether it’s “constructive” is hard to say. Certainly it looks much faster to compute (for general  $f$ ) than  $P$ , since  $|Y_n| \ll |\mathcal{P}_{P_n}|$ . Furthermore, though we couldn’t easily guess  $P$  by “statistical sampling” (since so many terms in its sum are 0), this may be no longer true for  $P'$ . So even though  $|Y_n|$  might well be too large for a

---

<sup>60</sup> For convenience, we’ll treat any boolean property or predicate of  $f$  interchangeably with the set of  $f$  for which it’s true.

<sup>61</sup> For a 1-output function, and using our simplest guess about pattern basis structure, that matrix has  $2^{2^1} = 4$  rows; a reasonable guess about a good choice of pattern basis for  $n = 1$  would result in some rows being trivial and the others being essentially equivalent, so a single row could be used. These details are illustrative; they don’t substantively affect the present discussion.

complete scan of its values to be fast enough, it doesn't seem safe in general to rule out  $P'$  being “constructive”.

But it's easy to see that  $P'$  is not “large” — at least if we manage to construct a pattern basis which fits the general description given in the Summary. Only a few very special functions will match a lot of the *same* patterns as our given functions  $\langle f_n \rangle$ ; even maximally complex functions, matching more patterns in  $\mathcal{P}_{P_n}$  than  $f_n$  does, would only extremely rarely have significant overlap with the patterns in  $Y_n$ . (This follows from our requirement that in a pattern basis we want to call “useful”, all pattern matrices should be sparse. We'll go through this a bit more quantitatively, below.)

So  $P$  is “large” but not “constructive”, whereas  $P'$  might be “constructive” but is not “large”. But we're not out of danger — what if there is some intermediate-sized property  $P''$  which is both?

To see that this won't happen, imagine gradually expanding the sets  $\langle Y_n \rangle$  (in whatever way we like) until their likely degree of overlap with the patterns in random functions starts to become significant, which means the modified property  $P''$  (based on these expanded  $\langle Y_n \rangle$ ) might hold for a “large” (i.e. non-negligible) fraction of  $f$ .

But the requirement in Section 2.3, that “probing randomly chosen patterns for polynomial time should not find a nonzero term (for a discriminating pattern) in a random  $f$ ”, can be applied just as well to probing  $f$  for the patterns in these expanded  $\langle Y_n \rangle$ ; it then basically says “if  $P''$  is constructive, it's not large”, since being “constructive” means  $P''$  doesn't have time to find any matches in  $f$  for discriminating patterns, but most patterns in  $Y_n$  will be discriminating (i.e. will apply to relatively few functions).

The above argument is not yet quantitative, but can easily be made so — for example, for suitable  $\alpha, \beta > 0$ , we could assume random functions typically match  $(2^{2^n})^\alpha$  (mostly discriminating) patterns, and define a “discriminating pattern” as one matched by at most  $(2^{2^n})^\beta$  functions.

This argument is also not formal — for example, it assumes the sets of patterns in random functions are “approximately random” (as subsets of all patterns) for purposes of estimating their degree of overlap, but never formalizes this (let alone proves it). But for the present purpose, that's ok — all we're arguing here is that the “pattern basis approach” is not obviously ruled out by the natural proofs barrier.

## 6.2 Does it “relativize” or “algebrize”?

Besides the natural proofs barrier, there are two other well-known general barriers to proving complexity lower bounds or complexity-class separations: “relativization”

[BGS75] and “algebrization” [AW09]. For any proposed approach to proving lower bounds, it’s worth considering whether it addresses them.

If the present approach could be realized (by finding a useful pattern basis), I am fairly confident a lower-bound proof making use of it would not “relativize” (i.e. still work for circuit families that can use an arbitrary oracle), since the set of generating circuits (and therefore their pattern matrices) is completely dependent on the possible primitive computations — it must be extended to include an “oracle gate” (which lets a circuit access the oracle). Thus every detail of an oracle, combined with every possible dual pattern (which corresponds to one row of the pattern matrix for the oracle gate), would affect the question of whether a given pattern basis has sparse pattern matrices for all generating circuits. In other words, the presence and detailed definition of an oracle would completely change the low-level model of computation, and thus the definition of “useful pattern basis”, and thus the required proof. (In principle this depends on the nature of the proof that the useful pattern basis exists, but clearly not *all* such proofs would relativize. In fact, it seems likely that for some oracles, no useful pattern basis exists; if true, this means *no* proof depending on the existence of such a basis could relativize.)

Based on the discussion in [AW09], I would also guess the present approach would not “algebrize”, for similar reasons. The authors of [AW09] state:

... arithmetization simply fails to “open the black box wide enough.” In a typical arithmetization proof, one starts with a polynomial-size Boolean formula  $\phi$ , and uses  $\phi$  to produce a low-degree polynomial  $p$ . But having done so, one then treats  $p$  as an arbitrary black-box function, subject only to the constraint that  $\deg(p)$  is small. Nowhere does one exploit the small size of  $\phi$ , except insofar as it lets one evaluate  $p$  in the first place. The message of this paper has been that, to make further progress, one will have to probe  $\phi$  in some “deeper” way.

(“Arithmetization” is a known proof technique able to evade the barriers discovered before algebrization.)

It seems clear that any presentation of an explicit pattern basis with sparse pattern matrices for all generating circuits does at least probe the *model of computation* (the set of generating circuits) in a deep way; adding an “algebraic oracle” would change that model as profoundly as adding a boolean oracle. And a lower-bound proof using a pattern basis would make direct use of a hypothetical circuit’s small size, to limit the number of nonzero (or non-“small”) terms in that circuit’s pattern matrix. So my guess is that this barrier is not an issue. (But my understanding of algebrization is very limited, so I would welcome a more informed opinion.)

## 6.3 the significance of linear superpositions of pattern maps

The “linearization” of the pattern map, when defining the pattern matrix  $\text{PM}_{\mathcal{P}}(f)$  of an arbitrary boolean function  $f$  (as described in Section 4.3), is straightforward, and adds no new information about  $f$ ; so it might seem pointless. But what it does is allow us to view dual patterns, patterns, and circuits’ effects on them as **linear superpositions** (over  $K$ ) of (respectively) truth tables, “primitive” truth table properties, and “primitive” pattern maps, in a space large enough to account separately for every possible property of truth tables (and with  $K$  being any field we find useful for describing these superpositions). So when trying to understand what kind of complex pattern, when passing through a simple circuit like an S-box (with many parallel wires adjoined), could split up into a sum of only a *few* other patterns, we’re free to consider *arbitrary sums* of “primitive” properties (or dually, sums of truth tables) whose components interfere in desirable ways. (The importance of ensuring that this sum has only a few terms was discussed at the end of the Motivation section.)

This possibility, of analyzing circuit effects or function patterns as superpositions, is a key new mathematical feature of the pattern basis approach. It opens up possible analogies to other areas of math and physics in which a similar approach has been useful for decomposing signals and systems with complex behavior into a superposition of components with simpler behavior, like fourier analysis, or the study of quantum mechanical “wavefunctions” understood as sums of eigenvectors of linear operators (this analogy was briefly discussed in Section 5.6). (Fourier analysis of boolean functions themselves — viewing their truth tables as vectors in  $\mathbb{R}^{2^n}$  or  $\mathbb{C}^{2^n}$  — is well-known and has proven useful (see e.g. [Odo14]), but this proposal would suggest applying it “one level of exponent higher”, over the property space, a vector space of dimension more like  $2^{2^n}$  or  $(2^n)!$  rather than  $2^n$ .)

## 6.4 even a nonconstructive proof would be significant

In Section 2.14, we gave simple conditions for a pattern basis to be *useful*.<sup>62</sup> These mostly amount to the sparseness of the pattern matrices of all generating circuits (when expressed using that basis), along with relative *non*-sparseness (i.e. a high value of  $C_{M,\mathcal{P}}(f)$ ) for a “sufficiently interesting” set of boolean functions  $f$ . The only informal parts of that are what counts as “relatively high” and “sufficiently interesting”.

In the best case, we’d prove  $C_{M,\mathcal{P}}(f_n)$  is superpolynomial in  $n$  for at least one

---

<sup>62</sup> See also the proof outline in Section 2.15, formalized in Section 4.6.

explicit function family  $\langle f_n \rangle$  provably in NP (since this would immediately give us  $\text{NP} \not\subseteq \text{P/poly}$ ).<sup>63</sup> But even a *superlinear* lower bound for an explicit family in NP would be a breakthrough.

Any such result would require understanding the pattern basis well enough to prove that many specific patterns matched the functions in  $\langle f_n \rangle$ . But it’s also possible we could find what *seems* like a useful pattern basis  $\mathcal{P}$  (since its pattern matrices are sparse enough), but be unable to prove any bound for an explicit function family (for example, if the patterns were mostly “too complicated to understand”). Maybe the only evidence for  $\mathcal{P}$ ’s “usefulness” would be a nonconstructive proof about “most  $f$ ”, like we have now for actual complexity.

But it’s important to understand that even that situation would be profoundly different from the one we’re in now.

In the present situation, we already have a simple counting argument which shows us that for *most*  $f$  (in a typical circuit model for general boolean circuits),  $C_{\min}(f)$  is high (relative to the maximum  $C_{\text{cost}}(g)$  over the generating circuits  $\mathcal{G}$ ) — in other words, that it’s easy in principle to find specific  $f$  for which the *actual* circuit complexity is high.

Unlike with Kolmogorov complexity (which can’t be proven high for *any* specific  $f$ , even in principle<sup>64</sup>), even at present a high  $C_{\min}(f)$  *can* be proven for specific  $f$  — though *only* in principle (when restricted to currently understood kinds of proofs), since the only known proofs would simply examine every possible circuit smaller than a specific size, and show that it computes some other function than  $f$ .

But even if we could somehow see and verify such a proof right now, we’d learn nothing interesting — such a proof would just brute-force the problem, encoding no understanding of it. In contrast, a proof of high  $C_{\min}(f)$  involving a useful measured complexity function  $C_{M,\mathcal{P}}$ , even for a generic but specific  $f$  given only by its truth table, would work by (effectively) computing and summing a large number of non-negative terms (the magnitudes of matrix entries of  $\text{PM}_{\mathcal{P}}(f)$ ), each based on some

---

<sup>63</sup> In an even better case, if we proved it *exponential* in  $n$  (as it presumably would be for the best choice of pattern basis  $\mathcal{P}$ , if this proposal works as intended), we’d establish the Exponential Time Hypothesis [IP01]. (See the Introduction for other references relevant to this subsection.)

<sup>64</sup> Proof by contradiction: if we could prove some explicit  $f$  had high Kolmogorov complexity, then the pseudocode “find the first such proof in which such an  $f$  appears in a specified explicit form” could be expanded to give us a short program to output such an  $f$  (so at least *that*  $f$  would have low  $C_K$ ). This is essentially the original proof from [Cha71] of what’s now called “Chaitin’s Incompleteness Theorem”. Formally, this must be parametrized by both a formal proof system and a universal Turing machine, but for reasonable choices, the absolute bound on provable Kolmogorov complexity is surprisingly small (see <http://web.archive.org/web/20131029224743/http://cs.umaine.edu/~chaitin/xgodel2.1>).

property about  $f$  not shared (for most of those properties) by most functions. But unlike the properties used by the brute-force proof (“ $f$  doesn’t match the function computed by *this* circuit”), these properties would multiply and combine in an organized way when functions were composed (since they’re organized as the entries of matrices being multiplied), so at least to that extent they’d have an understandable meaning and role in the proof.

Of course such a proof would still be only a thought experiment, since it would probably be no shorter than the brute-force kind when applied to a truly generic function  $f$  — such a function is too complex even to write down, except for very small values of  $n$ ; and the proof (in this generic- $f$  case) still has to be exponentially longer than the written form of  $f$ , to iterate through all the patterns in a pattern basis, or even just the ones with high pattern values for this  $f$ .

But even so, and even if an  $f$  for which this worked was only established nonconstructively, knowing that this kind of proof existed (especially if we knew this was true for most  $f$ ) would change our understanding of complexity; it might well give us new insight into its nature, and allow further consequences to be derived.

That said, our true goal goes further — we want to be able to prove high complexity for *interesting* (and easily describable) specific functions  $f$ , and to create such proofs in practice in an understandable way. We discuss in other sections why we hope this framework can help lead to both of these goals.

## 6.5 Conclusion

In a sense, this whole proposal can be considered as just a logical consequence of “taking the idea of hidden patterns seriously”. That is, it’s clear from experience that we humans *can* see or understand certain patterns, which *do* have significant effect on the behavior of systems they’re part of; but it’s also clear that there are significant patterns we aren’t able to see (for example, those latent in encrypted data, or in physically mixed systems which could be unmixed if time could be reversed).

The present proposal suggests that the patterns we can see or explicitly describe are only the simplest of a much larger set of patterns, most of which are unimaginably complex; yet all of them have significance in what happens (which makes them very special within the much *larger* set of “all possible properties”).

But this proposal also comes with the hope that the whole scheme of patterns can in some sense be understood, even though it’s hopeless to understand most patterns individually — indeed most of them are so complex (or to put it another way, they’re so numerous as a class) that we’ll never encounter more than a tiny fraction of them in any individual way (such as write them down or think about them).



But much like statistical mechanics can draw definite conclusions from considering the collective effect of innumerable detailed arrangements and motions of molecules, without ever attempting to describe each possible arrangement, we’re suggesting it might be possible to do something similar with all the patterns in a function which affect how it can be computed. In fact, by taking *all* of them into account (i.e. by thinking about the linear space of *all possible properties* of a function, and requiring that our pattern basis is sufficient to span it), we end up being able to see them as much more “controlled”, and in principle predictable, than otherwise (since function composition can then be seen as a bilinear operation on the set of patterns in each function).

## Acknowledgements

This paper wouldn’t exist without the excellent expository blogs of Scott Aaronson and Timothy Gowers, which introduced me to the status of circuit complexity and to the “natural proofs barrier”. More generally, I’m grateful to many authors for making readily available online their blog posts, preprints, course notes, and Wikipedia edits; as an amateur mathematician, it would not otherwise be practical to become informed about new fields.

I also thank Scott Aaronson for valuable feedback and suggestions, and John Baez for useful discussions and helping clarify some of my notation and terminology. (Any errors are of course my own.)

## References

- [–] (Note: references to blog comments are labelled with my own summaries.)
- [Aar07] Scott Aaronson, blog comment [conjectured pseudorandomness in  $\mathbb{F}_2$ -linear circuits] (2007) <http://www.scottaaronson.com/blog/?p=272#comment-7622>
- [Aar07a] Scott Aaronson, blog comment [superlinear lower bounds for an explicit  $\mathbb{F}_2$ -linear function would be a breakthrough] (2007) <http://www.scottaaronson.com/blog/?p=272#comment-7626>
- [Aar16] Scott Aaronson,  $P \stackrel{?}{=} NP$  [to appear] (2016)
- [AW09] Scott Aaronson and Avi Wigderson, *Algebrization: A new barrier in complexity theory*, ACM Trans. Comput. Theory, 1(1) (2009) 1–54

- [All96] Eric Allender, *Circuit complexity before the dawn of the new millennium*, Conference on Foundations of Software Technology and Theoretical Computer Science (FST&TCS), vol. 1180 of Lecture Notes in Computer Science (1996) 1–18
- [All08] Eric Allender, *Cracks in the defenses: Scouting out approaches on circuit lower bounds*, Computer Science — Theory and Applications. Springer Berlin Heidelberg, 2008. 3–10.
- [AB09] Sanjeev Arora and Boaz Barak, *Computational Complexity: A Modern Approach*, Cambridge University Press, 2009 (draft at <http://www.cs.princeton.edu/theory/complexity>)
- [BGS75] Theodore Baker, John Gill, and Robert Solovay, *Relativizations of the  $P = ? NP$  question*, SIAM Journal on Computing, 4 (4) (1975) 431–442
- [Ben73] C. H. Bennett, *Logical Reversibility of Computation*, IBM Journal of Research and Development, 17 (6) (1973) 525–532
- [BPP00] Joan Boyar, René Peralta, and Denis Pochuev, *On the multiplicative complexity of Boolean functions over the basis  $(\wedge, \oplus, 1)$* , Theoretical Computer Science 235 (2000) 43–57
- [Cha71] Gregory J. Chaitin, *Computational complexity and Gödel’s incompleteness theorem*, ACM SIGACT News 9 (1971), 11–12
- [Coh00] Bram Cohen, *The Unrelated Xors problem* (2000) (archived on Dec 21, 2001 at [https://web.archive.org/web/20011221212949/http://bitconjurer.org/unrelated\\_xors.html](https://web.archive.org/web/20011221212949/http://bitconjurer.org/unrelated_xors.html)) (see also Scott Aaronson’s comments on this conjecture, at <http://www.scottaaronson.com/blog/?p=272#comment-7622> and <http://www.scottaaronson.com/blog/?p=272#comment-7623>)
- [CG75] Don Coppersmith and Edna Grossman, *Generators for Certain Alternating Groups with Applications to Cryptography*, SIAM Journal on Applied Mathematics 29 (4) (1975) 624–627
- [For09] Lance Fortnow, *The status of the  $P$  versus  $NP$  problem*, Communications of the ACM 52 (9) (2009) 78–86
- [GGM86] Oded Goldreich, Shafi Goldwasser, and Silvio Micali, *How to construct random functions*, Journal of the ACM 33 (4) (1986) 792–807

- [Gow01] Timothy Gowers, *A new proof of Szemerédi’s theorem*, Geometric and Functional Analysis 11 (3) (2001) 465–588
- [Gow09] Timothy Gowers, *A conversation about complexity lower bounds, continued*, Gowers’s Weblog (2009), <https://gowers.wordpress.com/2009/09/28/a-conversation-about-complexity-lower-bounds-continued>
- [Gow09a] Timothy Gowers, *A conversation about complexity lower bounds, VIII*, Gowers’s Weblog (2009), <https://gowers.wordpress.com/2009/10/27/a-conversation-about-complexity-lower-bounds-viii/>
- [IP01] Russell Impagliazzo and Ramamohan Paturi, *On the Complexity of  $k$ -SAT*, Journal of Computer and System Sciences 62(2) (2001) 367–375
- [KKL88] Jeff Kahn, Gil Kalai, and Nathan Linial, *The Influence of Variables on Boolean Functions*, Proceedings of the 29th Symposium on the Foundations of Computer Science (1988) 68–80
- [MB99] Saunders Mac Lane and Garrett Birkhoff, *Algebra (third edition)*, AMS Chelsea Publishing (1999)
- [MS01] Ketan D. Mulmuley and Milind Sohoni, *Geometric complexity theory I: An approach to the  $P$  vs.  $NP$  and related problems*, SIAM Journal on Computing 31, 2 (2001) 496–526
- [Mul12] Ketan D. Mulmuley, *The GCT program toward the  $P$  vs.  $NP$  problem*, CACM 55, 6 (2012) 98–107
- [NC00] Michael Nielsen and Isaac Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press (2000)
- [Odo14] Ryan O’Donnell, *Analysis of Boolean Functions*, Cambridge University Press (2014)
- [PW10] Mihai Pătraşcu and Ryan Williams, *On the possibility of faster SAT algorithms*, Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms (SODA ’10) (2010) 1065–1075
- [RR97] Alexander Razborov and Steven Rudich, *Natural proofs*, Journal of Computer and System Sciences, 55(1):24–35, 1997
- [Reg02] Kenneth W. Regan, *Understanding the Mulmuley-Sohoni Approach to  $P$  vs.  $NP$* , Bulletin of the EATCS 01 (2002) 78:86–99

- [Seg14] Ed Segal, *lecture notes on Group Representation Theory*, (2014) (available from <http://wwwf.imperial.ac.uk/~epsegal/repthy.html>)
- [She08] Alexander A. Sherstov, *The pattern matrix method for lower bounds on quantum communication*, Proc. of the 40th Symposium on Theory of Computing (STOC) (2008) 85–94
- [SM02] Larry Stockmeyer and Albert R. Meyer, *Cosmological Lower Bound on the Circuit Complexity of a Small Problem in Logic*, Journal of the ACM 49 (6) (2002) 753–784
- [TZ11] Terence Tao and Tamar Ziegler, *The Inverse Conjecture for the Gowers Norm over Finite Fields in Low Characteristic*, Annals of Combinatorics 16: 121–188, 2011
- [Tre08] Luca Trevisan, blog comment [about a typical fallacy in claimed proofs of  $P \neq NP$ ] (2008) <http://www.scottaaronson.com/blog/?p=304#comment-8984>
- [VO05] A. M. Vershik and A. Yu. Okounkov, *A New Approach to the Representation Theory of the Symmetric Groups. II*, (2005) arXiv:math/0503040
- [Web16] Peter Webb, *A Course in Finite Group Representation Theory*, Cambridge University Press (2016) (draft at <http://www.math.umn.edu/~webb/RepBook/>)
- [WikF] Wikipedia contributors, *Factor analysis* (retrieved April 19, 2016 from [https://en.wikipedia.org/w/index.php?title=Factor\\_analysis&oldid=715136055](https://en.wikipedia.org/w/index.php?title=Factor_analysis&oldid=715136055))
- [WikG] Wikipedia contributors, *Gowers norm* (retrieved March 31, 2016 from [https://en.wikipedia.org/w/index.php?title=Gowers\\_norm&oldid=697660135](https://en.wikipedia.org/w/index.php?title=Gowers_norm&oldid=697660135))
- [WikO] Wikipedia contributors, *Big O notation* (retrieved March 25, 2016 from [https://en.wikipedia.org/w/index.php?title=Big\\_O\\_notation&oldid=711907606](https://en.wikipedia.org/w/index.php?title=Big_O_notation&oldid=711907606))
- [WikR] Wikipedia contributors, *Representation theory of finite groups* (retrieved April 20, 2016 from [https://en.wikipedia.org/w/index.php?title=Representation\\_theory\\_of\\_finite\\_groups&oldid=716249531](https://en.wikipedia.org/w/index.php?title=Representation_theory_of_finite_groups&oldid=716249531))
- [Wil11] Ryan Williams, *Week 1: An Overview of Circuit Complexity*, class notes for CS354 (2011) (retrieved from <http://web.stanford.edu/~rrwill/week1.pdf>)

- [Wil14] Ryan Williams, *Algorithms for Circuits and Circuits for Algorithms: Connecting the Tractable and Intractable*, Proceedings of the International Congress of Mathematicians (2014) (draft at <http://web.stanford.edu/~rrwill/ICM-survey.pdf>)
- [Zag04] Don Zagier, *Applications of the representation theory of finite groups*, Appendix to S. Lando and A. Zvonkin, *Graphs on Surfaces and Their Applications*, Encycl. of Math. Sciences 141, Springer-Verlag, Berlin Heidelberg (2004), 399–427 (retrieved from <http://people.mpim-bonn.mpg.de/zagier/files/tex/ApplRepTheoryFiniteGroups/fulltext.pdf>)